

On the Privacy Implications of Real Time Bidding

Muhammad Ahmad Bashir
PhD Thesis Defense
August 13, 2019

Committee:

Christo Wilson (Advisor)
David Choffnes
William Robertson
Arvind Narayanan (Princeton)

Northeastern University
Khoury College of
Computer Sciences

User Concern Regarding Online Privacy

User Concern Regarding Online Privacy

News | Opinion | Sport | Culture | Lifestyle



The
Cambridge
Analytica
Files

Cambridge
Analytica

This article is more than 1 year old

Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

BUSINESS
INSIDER

TECH | FINANCE | POLITICS | STRATEGY | LIFE | ALL

BI PRIME | INTELLIGENCE



Whistleblower describes how firm linked to former T. Steve Bannon compiled user data to target American

'I made Steve Bannon's psychological warfare tool' data war whistleblower

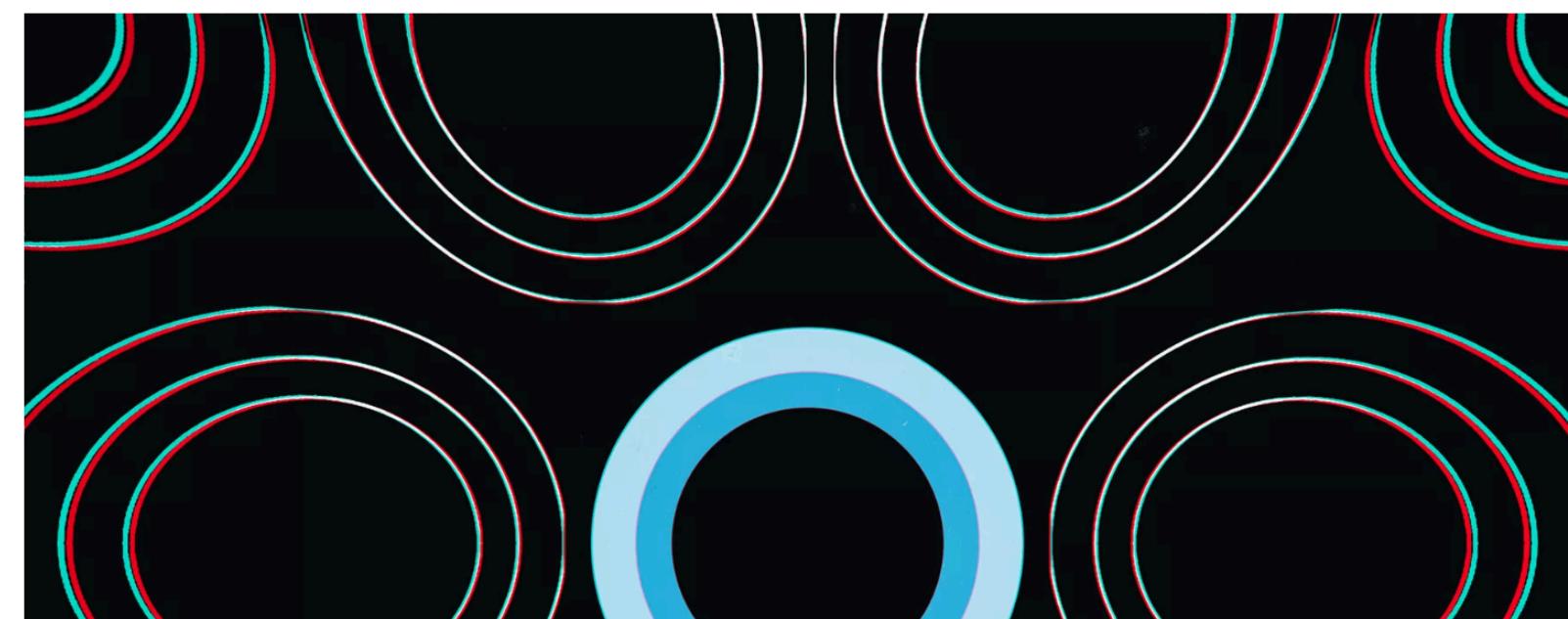
MICROSOFT | POLICY | TECH

Microsoft contractors are listening to select Skype calls and Cortana recordings

Microsoft does not explicitly say humans listen to the recordings

By Jon Porter | @JonPorty | Aug 7, 2019, 11:14am EDT

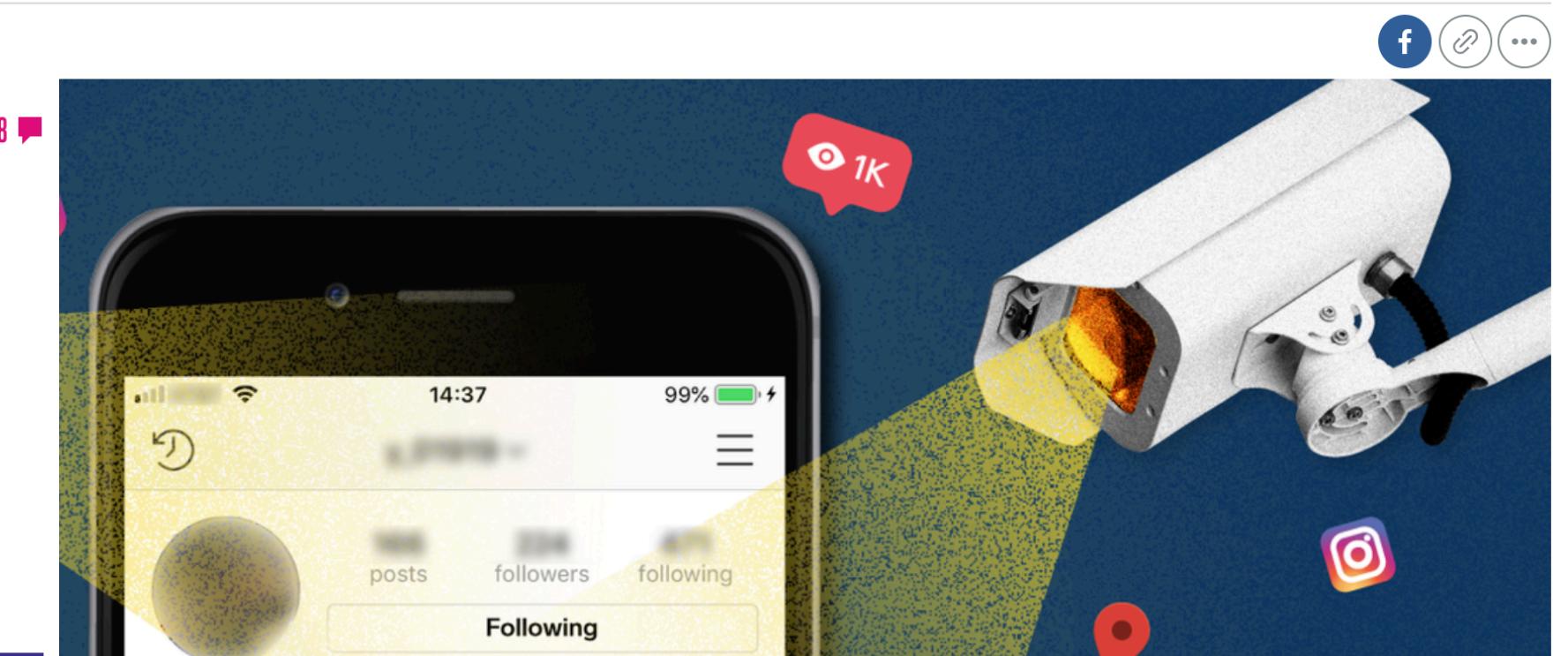
f t SHARE



GOOD DEALS



At \$200 off, the latest MacBook Air is an even better deal at Best Buy



Privacy on the Web

FOOTBALL NEWS

LATEST NEWS



Advertisement

TODAY'S MATCHES

TRANSFERS 18:51

We use cookies to give you the best online experience. By using our website you agree to our use of cookies in accordance with our privacy policy.

SHOW MORE

ACCEPT

Privacy on the Web

FOOTBALL NEWS

LATEST NEWS

TODAY'S MATCHES



Tottenham agree €70m Dybala fee with Juventus

Advertisement

TRANSFERS 18:51

We use cookies to give you the best online experience. By using our website you agree to our use of cookies in accordance with our privacy policy.

SHOW MORE | ACCEPT

A black arrow points from the word "Advertisement" in the top right section towards the bottom right corner of the page, where a red border highlights a cookie consent banner.

Privacy on the Web

FOOTBALL NEWS

LATEST NEWS

TODAY'S MATCHES



Tottenham agree €70m Dybala fee with Juventus

Advertisement

TRANSFERS 18:51

We use cookies to give you the best online experience. By using our website you agree to our use of cookies in accordance with our privacy policy.

SHOW MORE | ACCEPT

A black arrow points from the word "Advertisement" in the top right section towards the bottom right corner of the page, where the "ACCEPT" button is located.

Privacy on the Web

Perform Media - Privacy Notice

Perform Media - Privacy Notice

1. Introduction

Perform Media ("we", "us" or "Perform Media") is a business division of the Perform Group, headquartered in the UK. Perform Media is a group of companies within the Perform Group which own, operate (either for ourselves or on behalf of third-parties) or represent a number of free-to-use digital sports media properties, which are either websites or mobile applications. We also operate a video-content streaming service called ePlayer, which allows third-party digital publishers to stream sports video content on their own digital platforms. In this notice, each of the properties and ePlayer platforms is referred to as a "platform". When we refer to the "Service" in this Privacy Notice, this means the provision of the content on the platforms to the user.

This notice explains why Perform Media collects information concerning the users of the platforms (each a "user", "you" or "your"), what type of information we collect and how we use that information. This Privacy Notice applies to any such information collected via any platform which displays it (or a link to it).

Additional information for users located in the European Economic Area ("EEA"), Switzerland and the UK about their rights and other information can be found in the Annex to this Privacy Notice.

This notice has been last updated on 25 May 2018.

2. What data we collect

When a user interacts with any of the platforms, Perform Media may collect certain data about that user. This data falls into 2 broad categories: (i) data about a user themselves, which the user gives us; and (ii) data about the user, or the way a user interacts with the platforms or Services, which is collected by us electronically (collectively "User Information").

We collect the following User Information in accordance with applicable law:

- information that users provide to us, directly or via a third-party registration service for the purpose of registering with us or our group companies;
- a user's entry details for a competition or promotion;
- information users provide to us when users interact with our customer service team;
- information users provide when users participate in our surveys or marketing promotions; and
- information from a user's social media profile to populate their profile on our platforms, and where the users chooses to register with us using their social media login(s);
- information users provide relating to reviews or ratings, favourite sports, tournaments, teams, events or personalities, and/or other subjects of interest to sports fans.
- information about user interactions with our platforms, and any direct communications the user receives from us (such as emails);
- device IDs or unique identifiers, location data, device and software characteristics (such as type and configuration), connection information, statistics on page views, referral URLs, ad data, website and app navigation, IP address and standard web log information;
- the language choices made by users in relation to their use of the Service, which may include information relating to their location when using the Service;
- the details of any third-party digital property (such as a website) that referred or linked a user to the platforms, together with information about how a user interacted with any online and email marketing campaigns related to the Service;
- information and communications on forums on the platforms, including chat rooms and message boards;
- information about how users behave on our platforms (such as the type of content they view, how frequently, at what times etc.).

We may also collect personal information from a variety of third-party sources. These include social networks (subject to the user's privacy settings and the privacy policy of the social network, and adopting a "two-click" solution where required), service providers which allow us to verify, enhance or supplement the User Information collected on our platforms, and business partners to whom we are offering a service.

The platform may include Nielsen's digital measurement software, which allows users to contribute to market research. To learn more about the data collected by Nielsen's

software and your options you can visit Nielsen measurement Privacy Policy at this [link](#).

3. How & why we use the data we collect

We use the User Information for three main purposes: (i) to operate and improve the platforms and to ensure that we can provide the Services; (ii) to optimise the content on the platforms; and (iii) send marketing and show advertising to our users.

More specifically, we will use the User information to do one or more of the following things:

- a. administer our platforms;
- b. provide the Service requested by the user;
- c. improving the browsing experience by personalizing the Service to the user;
- d. recommend goods, services or promotions which may be of interest to the user or contact users in order to conduct market research (with the consent of the user, where necessary). For these purposes, we will create a profile about your interests and preferences which will enable us to increase our advertising and communication relevance;
- e. prevent, detect and investigate potentially prohibited or illegal activities, including fraud, cheating, and money laundering and enforce our terms;
- f. to notify users about changes to our Service;
- g. send out email alerts, text messages notifications or device notifications in relation to our Service;
- h. send users prizes won via the platforms, and supply services purchased via our platforms;
- i. send notices and invoices to users, and collect payments from them;
- j. provide third parties such as our third-party digital publishers who deploy ePlayer with statistical and other information about our users as required to operate the platforms or Services;
- k. deal with enquiries and complaints;
- l. provide our commercial partners with information about users' use of co-branded services, including historical use data;
- m. comply with any legal or regulatory requirements and otherwise to any respond to any relevant regulator or competent authority as required by law;
- n. understand user trends and patterns; and
- o. develop aggregate analysis in respect of our users and business intelligence that enable us to operate, protect, make informed decisions and report on the performance of our business.

The digital sports media platforms we operate are designed to attract as many users as possible. In our case, our users are typically sports fans or interested in sports news and other sports-related content. The creation of this content, and the work that goes into ensuring that the platforms continue to be class-leading digital sports destinations, are funded by advertising revenues which we receive as a result of showing users advertising or other sponsored or branded content on the platforms. We have a legitimate interest in ensuring that we deliver this advertising in a responsible way, and in allowing us to deliver it to the users most likely to be interested in it, and in doing so we employ User Information, such as the user's location, how the user engages with the platforms and the content that the user views.

More specifically, in relation to advertising:

- we employ User Information to select and deliver advertising that appears on our platforms, and to allow third-parties (such as ad servers) to do this on our behalf;
- we employ User Information to select and deliver advertising to users based on the way the user has interacted with our platforms, and to share that information with third-parties to allow them to do so;
- we may pass User Information to service providers, such as Oracle or Google, to allow them to provide services on our behalf or to partner with us in selecting and serving ads for our advertising partners;
- we may pass User Information to advertisers or third-party publishers who display the ePlayer, to allow us to report on the advertising displayed to our users and how users interacted with it;
- we will not employ User Information to make qualitative judgments about individual users, nor make automated decisions in relation to users which could have a significant or legal effect on them, nor will we knowingly engage any third-party to do so.

4. Cookies

Much of the User Information that we collect, and most of the information we employ in connection with advertising, is obtained via the use of cookies and similar technologies. Cookies are usually a string of numbers and/or letters that a website transfers to a user's hard drive. Similar technologies include pixel tags and web storage and other technologies, and we refer to all these things as "cookies" in this section. The cookies enable the

and we refer to all these things as "cookies" in this section. The cookies enable the platforms to "remember" you, either for the duration of your visit (session cookies) or for repeat visits (persistent cookies).

On this platform, we use the following types of cookies:

- **Strictly necessary cookies** - These cookies are essential to enable the user to navigate around the platform securely and to provide the user with Services they have specifically requested.
- **Functionality cookies** - These cookies enhance the functionality of the platforms by storing user preferences. For instance, they can remember your name and location, if you provide this information and the types of content you are interested in.
- **Performance cookies** - These cookies improve the performance of the platforms. For instance, they help pages load quicker.
- **Analytics cookies** - We may use third party analytics services such as Google Analytics and others. Your IP address and other information will be collected by automated means to evaluate your use of the website, compile reports on website activity, identify user patterns and provide related services. For more information about how Google Analytics uses your information please click [here](#). In addition, authentication and tracking logs will be used to compile user statistics.
- **Targeting/advertising** - These cookies help us serve of relevant advertisements and implement frequency capping procedures.

These types of cookies may also be placed on the platform by our business partners or service providers.

If you register with us or if you continue to use our platforms, you agree to our use of cookies. In some cases, when you arrive on our platform a pop-up message will appear asking for your consent to place advertising cookies on your device as required by applicable law. In order to provide your consent, please click 'OK'. Once your consent has been provided, this message will not appear again when you revisit, unless you delete cookies via your browser settings or otherwise.

Most web browsers automatically accept cookies, but if you prefer, you can change your browser to prevent that, or opt-out as explained below. You may not be able to take full advantage of our platforms or use certain functions if you disable cookies.

5. Disclosure of User Information

We may share User Information with any member of the Perform group (www.performgroup.com/) where this is necessary for business administration purposes in accordance with applicable law. We may also allow other members of the Perform group to contact users with relevant offers, competitions and latest news (where they have consented to doing so, if required by applicable laws).

We may also share User Information with selected third parties including:

- business partners, suppliers and sub-contractors in connection with providing Services to our users.
- advertisers and advertising networks that require the data to select and serve relevant adverts to the user.
- analytics and search engine providers that assist us in the improvement and optimisation of the platforms.
- if we decide to sell the business operating the Service or it is integrated with another business, we may disclose User Information to our advisers and any prospective buyer's advisers as well as to the new owners of the business.
- if we are under a duty to disclose or share a user's personal information in order to comply with any legal obligation, or in order to enforce or apply our terms;
- in order to prevent fraud, cheating and money laundering we may disclose User Information and betting and gaming history to third parties including, but not limited to, any relevant regulator, gaming and sporting bodies, financial institutions and law enforcement agencies or any other body that deals with the investigation of alleged offences.

Where relevant, we will disclose User Information and anonymized information associated with a user's use of and interaction with the platforms to the partner on whose behalf we operate that platform, for the purposes of marketing and advertising, and to promote their own products and services.

6. Your right to opt-out

If you would like us to stop sending you marketing communications you may use the opt-out link here, the unsubscribe link in our marketing communications or by amending your App settings. Although we encourage you to use the opt-out link and App settings because it is automated, you may also contact us [here](#).

If you, or another user of your device, wish to withdraw your cookies consent at any time, you have the ability to do so by deleting cookies from your browser settings. You

Privacy on the Web

5. Disclosure of User Information

We may share User Information with any member of the Perform group (www.performgroup.com/) where this is necessary for business

administration purposes in accordance with applicable law. We may also allow other members of the Perform group to contact users with relevant offers, competitions and latest news (where they have consented to doing so, if required by applicable laws).

We may also share User Information with selected third parties including:

- business partners, suppliers and sub-contractors in connection with providing Services to our users.
- advertisers and advertising networks that require the data to select and serve relevant adverts to the user.

Privacy on the Web

5. Disclosure of User Information

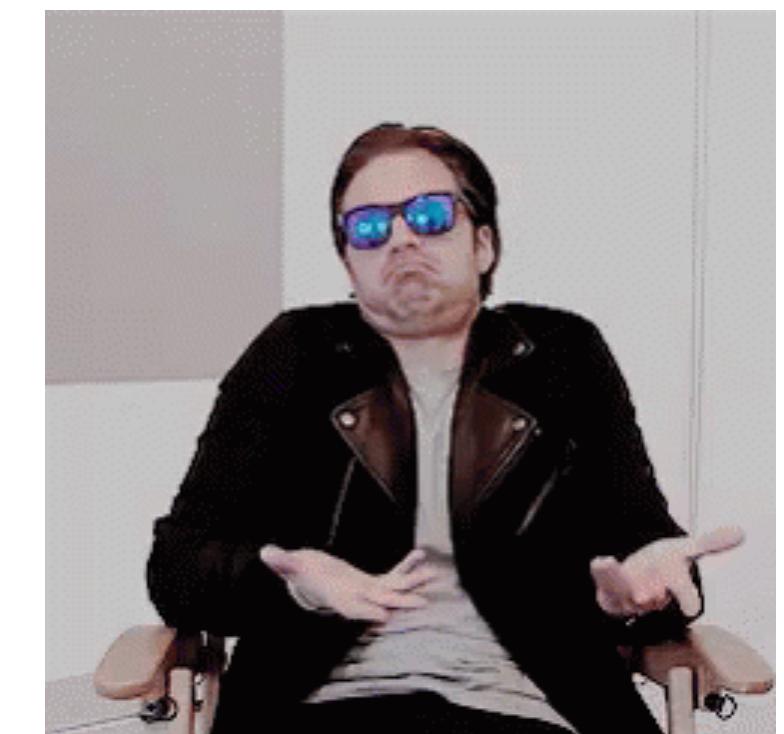
We may share User Information with any member of the Perform group (www.performgroup.com/) where this is necessary for business

administration purposes in accordance with applicable law. We may also allow other members of the Perform group to contact users with relevant offers, competitions and latest news (where they have consented to doing so, if required by applicable laws).

We may also share User Information with selected third parties including:

- business partners, suppliers and sub-contractors in connection with providing Services to our users.
- advertisers and advertising networks that require the data to select and serve relevant adverts to the user.

Who receives information?



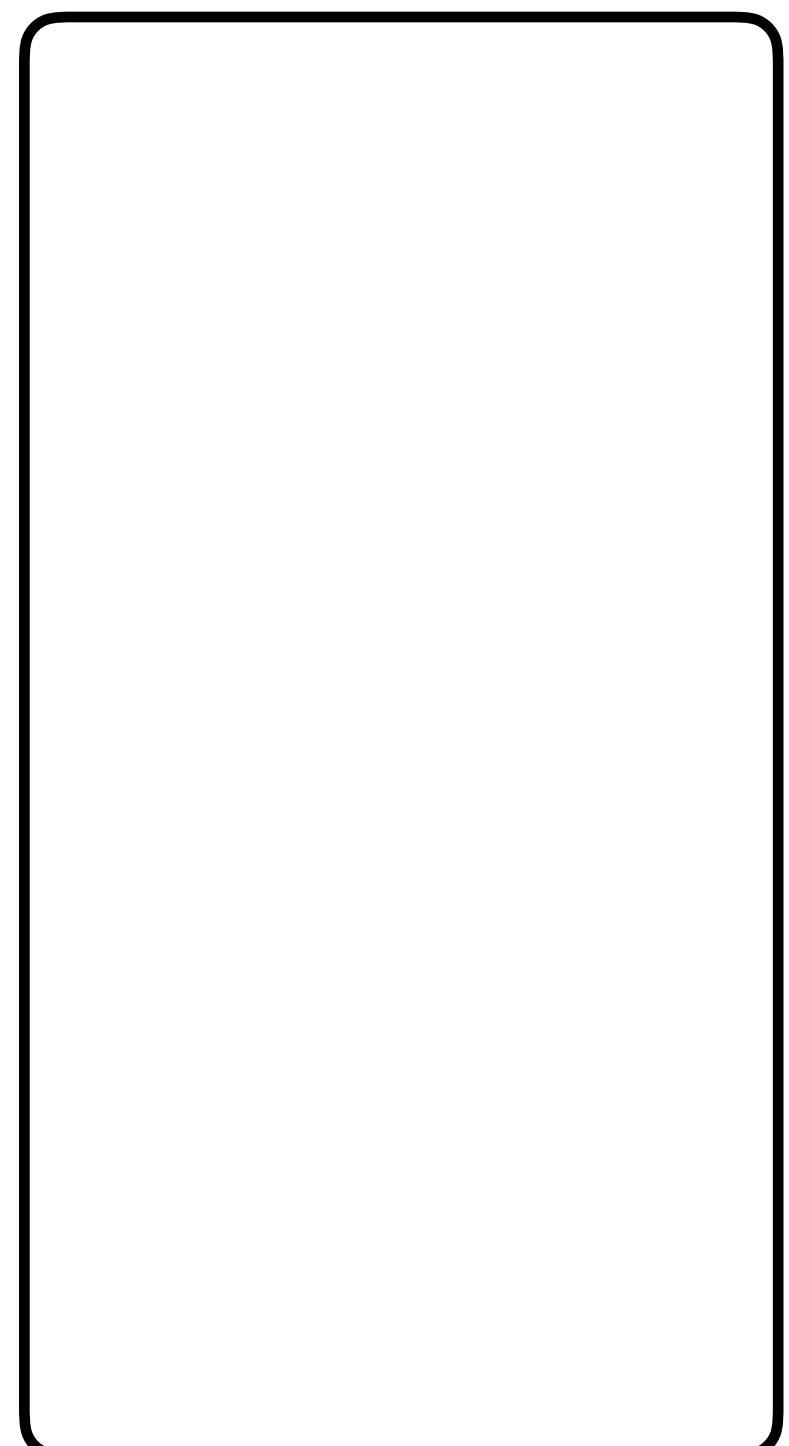
Privacy on the Web

The screenshot shows a web browser window for <https://www.dailymail.co.uk/home/index.html>. The page displays a 'How we personalise your experience' dialog. The 'Functional' tab is selected, while 'Advertising' is unselected. The dialog lists several third-party advertising companies with their respective privacy policies:

Company	Privacy policy
1&1 Internet	Privacy policy
1000mercis	Privacy policy
1000mercis	Privacy policy
1020, Inc. dba Placecast and Ericsson Emodo	Privacy policy
12 Digit Marketing	Privacy policy
12DigitMedia	Privacy policy
161Media	Privacy policy

At the bottom of the dialog are two buttons: 'Don't allow these partners' and 'Allow all'.

Who receives information?



Privacy on the Web

The screenshot shows a web browser window for <https://www.dailymail.co.uk/home/index.html>. At the top, there's a lock icon and the URL. Below the address bar, there's a "DVERTISEMENT" section with a "Like 16.1M" button and a weather forecast for Tuesday, Aug 6th 2019 (10PM 19°C, 1AM 18°C, 5-Day Forecast). The main content area has a large "How we personalise your experience" heading. It includes tabs for "Advertising" and "Functional". Under "Advertising", there's a list of third-party advertising companies: 1&1 Internet, 1000mercis, 1000mercis, 1020, Inc. dba Placecast and Ericsson Emodo, 12 Digit Marketing, 12DigitMedia, and 161Media. Each company name is followed by a "Privacy policy" link. A red box highlights the first two entries: 1&1 Internet and 1000mercis. At the bottom, there are "Don't allow these partners" and "Allow all" buttons.

Follow DailyMail breaking news and top stories from around the globe

Instagram

EXC anti by I

Like 16.1M

Tuesday, Aug 6th 2019 10PM 19°C 1AM 18°C 5-Day Forecast

How we personalise your experience

Advertising **Functional**

Third party advertising companies

Below is a list of the third-party advertising companies who may have access to this data or the purposes and features they each declare when you click "Got It" or "Allow All".

1&1 Internet	Privacy policy
1000mercis	Privacy policy
1000mercis	Privacy policy
1020, Inc. dba Placecast and Ericsson Emodo	Privacy policy
12 Digit Marketing	Privacy policy
12DigitMedia	Privacy policy
161Media	Privacy policy

[Don't allow these partners](#) [Allow all](#)

Who receives information?

1&1 Internet
1000mercis
Ericsson Emodo
Digit Marketing

Privacy on the Web

The screenshot shows a consent dialog from the Daily Mail website. At the top, there's a header with a lock icon, the URL <https://www.dailymail.co.uk/home/index.html>, and a like count of 16.1M. Below the header, there's a navigation bar with links like "Home", "Latest", and "Log in". The main content area has a title "How we personalise your experience" with tabs for "Advertising" and "Functional". Under the "Advertising" tab, there's a section titled "Third party advertising companies" with a list of companies: 1&1 Internet, 1000mercis, 1020, Inc. dba Placecast and Ericsson Emodo, 12 Digit Marketing, 12DigitMedia, and 161Media. Each company name is followed by a link to its privacy policy. The entire list of companies and their privacy policy links are highlighted with a red border. At the bottom of the dialog are two buttons: "Don't allow these partners" and a large green "Allow all" button.

Follow
DailyMail
breaking
s and top
ries from
ound the
globe

Like 16.1M

Tuesday, Aug 6th 2019 10PM 19°C 1AM 18°C 5-Day Forecast

Home Latest Log in

How we personalise your experience

Advertising Functional

Third party advertising companies

Below is a list of the third-party advertising companies who may have access to this data or the purposes and features they each declare when you click "Got It" or "Allow All".

1&1 Internet

1000mercis

1000mercis

1020, Inc. dba Placecast and Ericsson Emodo

12 Digit Marketing

12DigitMedia

161Media

Privacy policy

Don't allow these partners

Allow all

Who receives information?

1&1 Internet
1000mercis
Ericsson Emodo
Digit Marketing



Privacy on the Web

The screenshot shows a consent dialog from the Daily Mail website. At the top, there's a header with a lock icon, the URL <https://www.dailymail.co.uk/home/index.html>, and a date/time stamp: Tuesday, Aug 6th 2019 10PM 19°C 1AM 18°C 5-Day Forecast. Below the header, there's a sidebar with social media links for Follow DailyMail, Home, Latest, Instagram, and a large '10' icon. The main content area has a title 'How we personalise your experience' with tabs for 'Advertising' and 'Functional'. A green callout bubble points to a button labeled 'To facilitate Real Time Bidding (RTB)'. Below the tabs, under 'Advertising', is a section titled 'Third party advertising companies' listing partners: 1&1 Internet, 1000mercis, 1020, Inc. dba Placecast and Ericsson Emodo, 12 Digit Marketing, 12DigitMedia, and 161Media. To the right, under 'Functional', are multiple 'Privacy policy' links. At the bottom are two buttons: 'Don't allow these partners' and a large green 'Allow all' button.

Follow DailyMail breaking news and top stories from around the globe

Instagram

10

Like 16.1M

Tuesday, Aug 6th 2019 10PM 19°C 1AM 18°C 5-Day Forecast

How we personalise your experience

Advertising

Functional

Third party advertising companies

Below is a list of the third-party advertising companies that we work with to facilitate Real Time Bidding (RTB). You can choose to opt-out of this data or the purposes and features they provide by clicking "Got It" or "Allow All".

1&1 Internet

1000mercis

1000mercis

1020, Inc. dba Placecast and Ericsson Emodo

12 Digit Marketing

12DigitMedia

161Media

To facilitate Real Time Bidding (RTB)

Privacy policy

Don't allow these partners

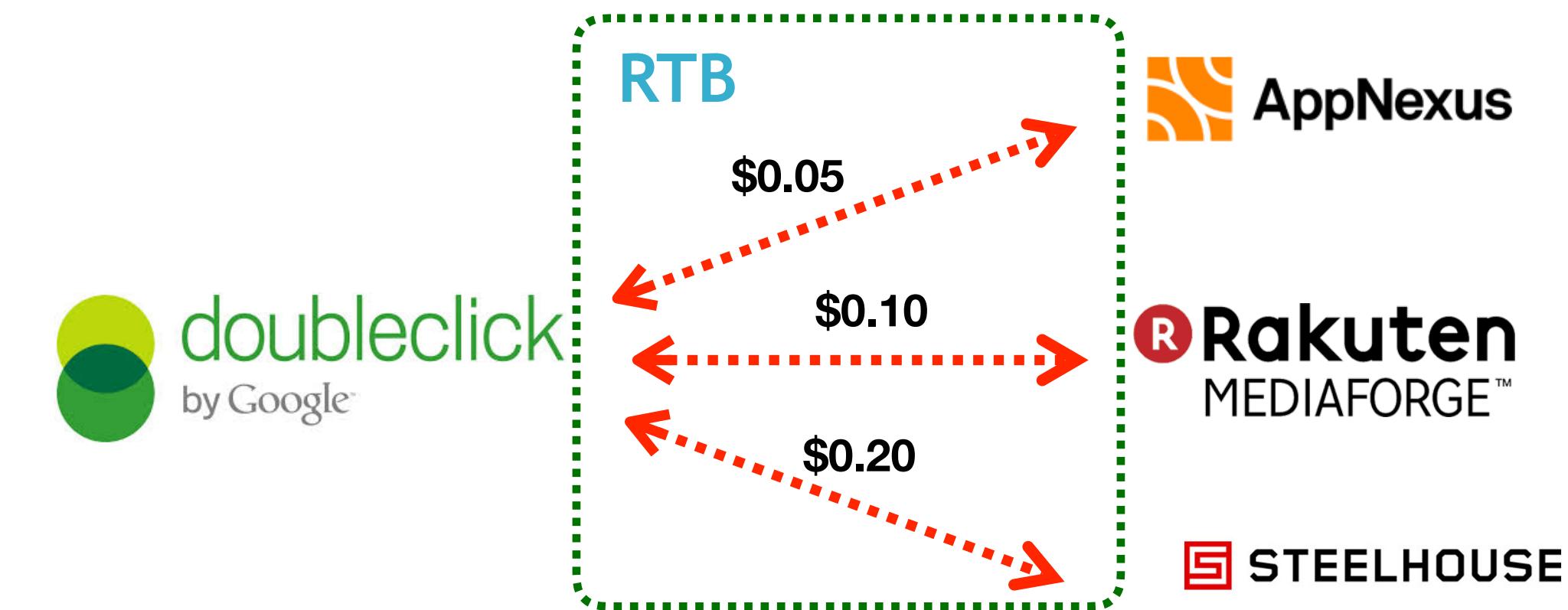
Allow all

Who receives information?

1&1 Internet
1000mercis
Ericsson Emodo
Digit Marketing



Thesis



Real Time Bidding (RTB) has increased collaboration among Advertising & Analytics (A&A) companies, which increases privacy exposure for end-users

Companies Track You

Publishers



Advertising & Analytics (A&A) companies

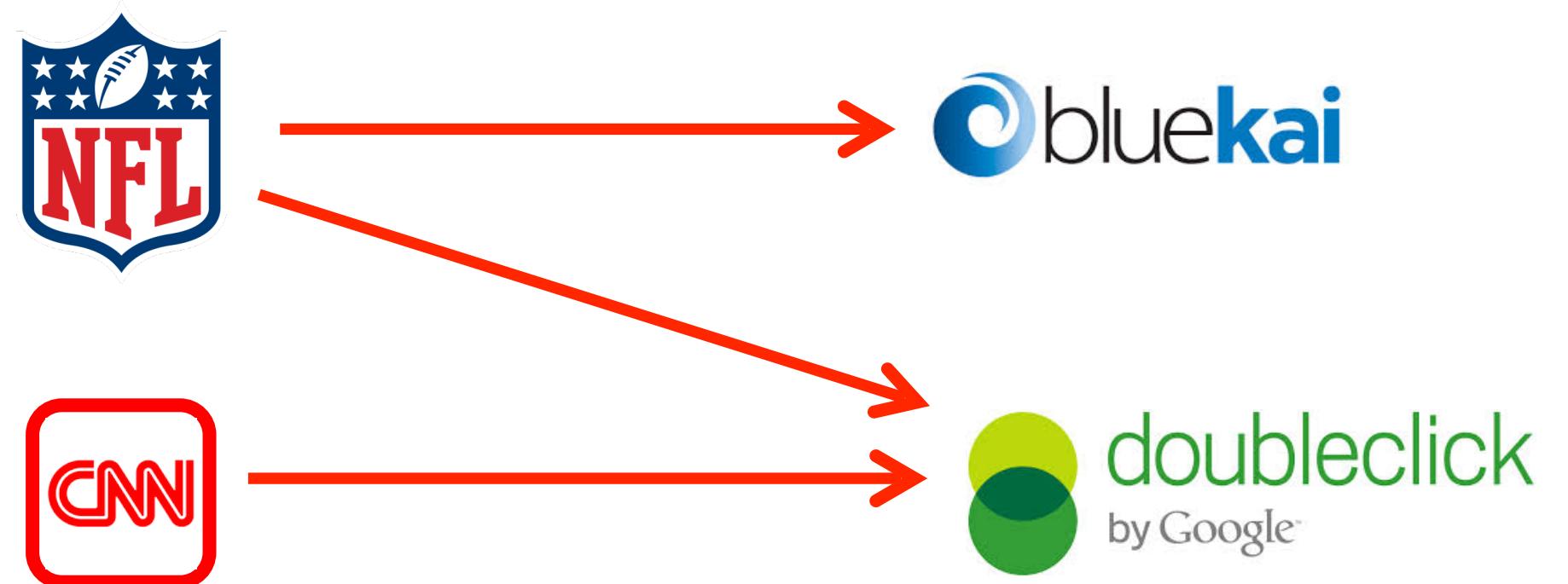
→ Observable via browser

Companies Track You

Publishers

Advertising & Analytics (A&A) companies

Observable via browser

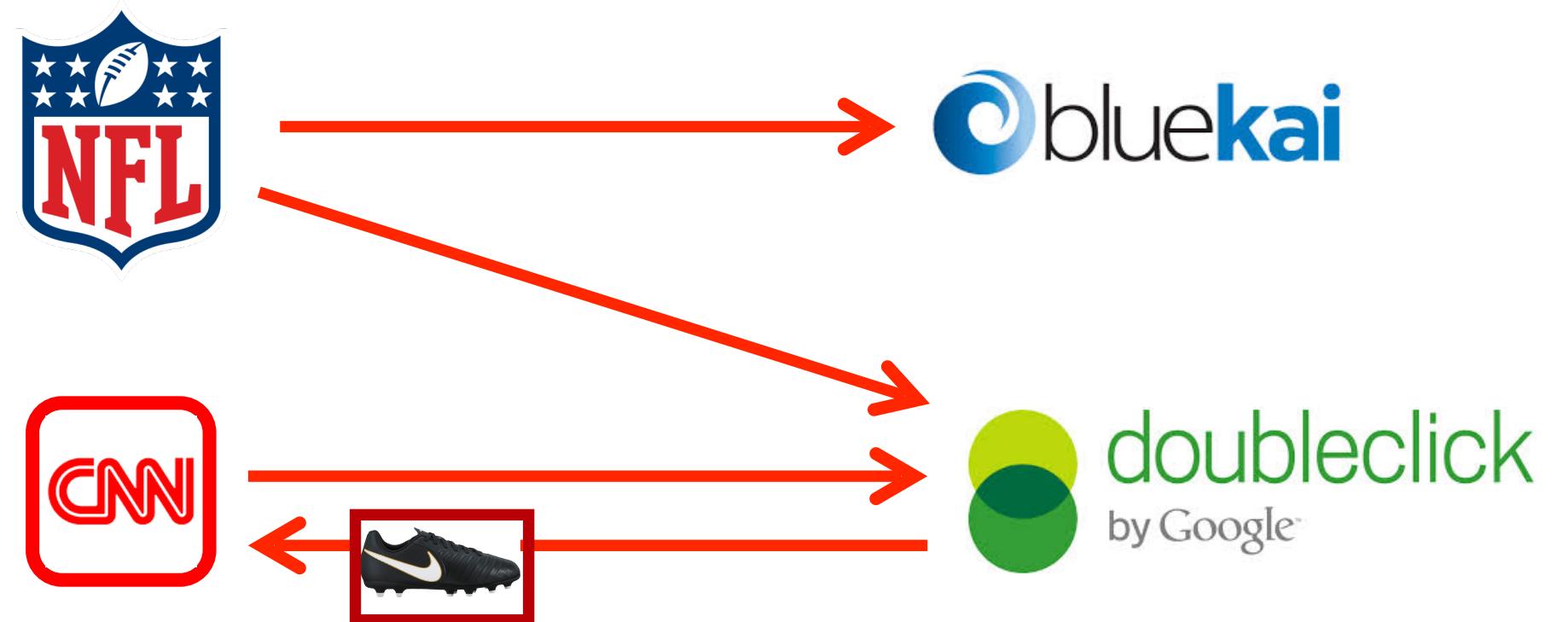


Companies Track You

Publishers

Advertising & Analytics (A&A) companies

→ Observable via browser

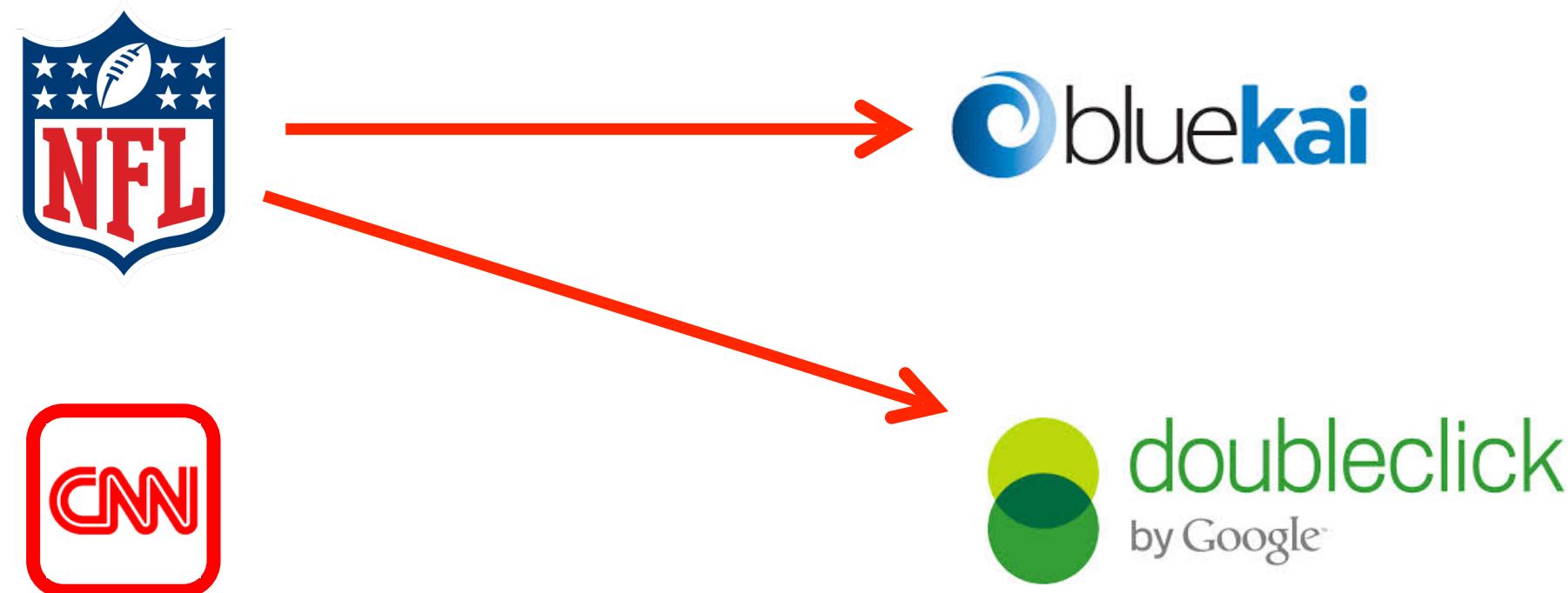


Ecosystem is Moving Toward Real Time Bidding (RTB)

Publishers

Advertising & Analytics (A&A) companies

→ Observable via browser

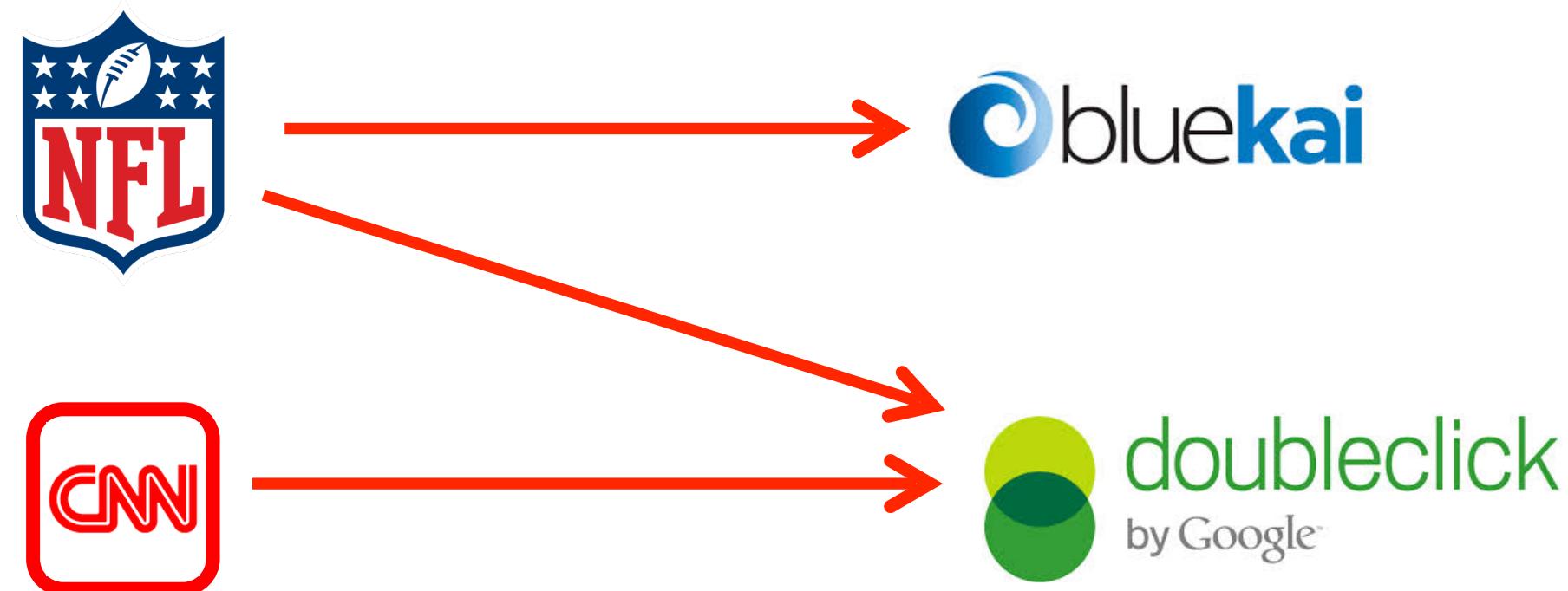


Ecosystem is Moving Toward Real Time Bidding (RTB)

Publishers

Advertising & Analytics (A&A) companies

→ Observable via browser



Ecosystem is Moving Toward Real Time Bidding (RTB)

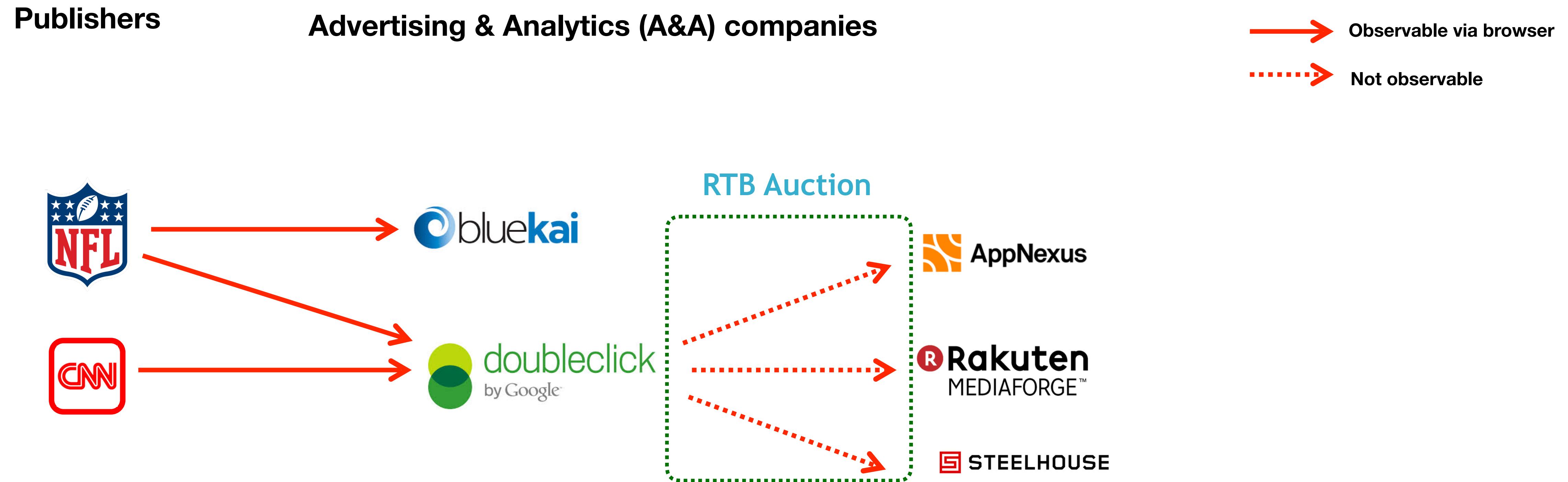
Publishers

Advertising & Analytics (A&A) companies

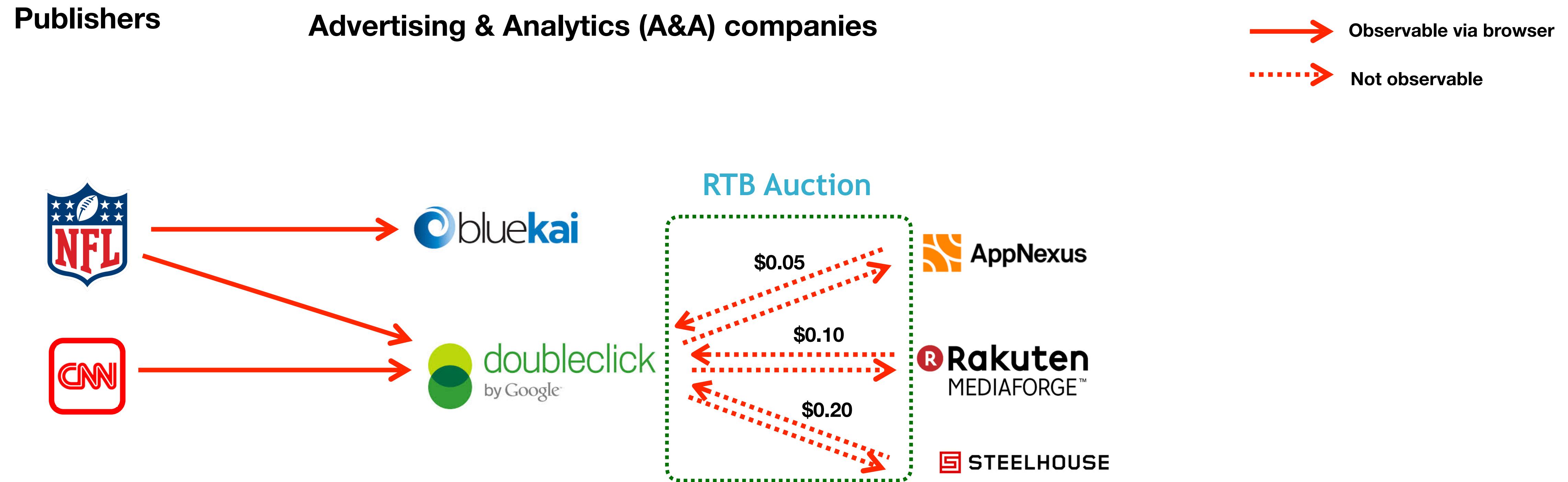
→ Observable via browser



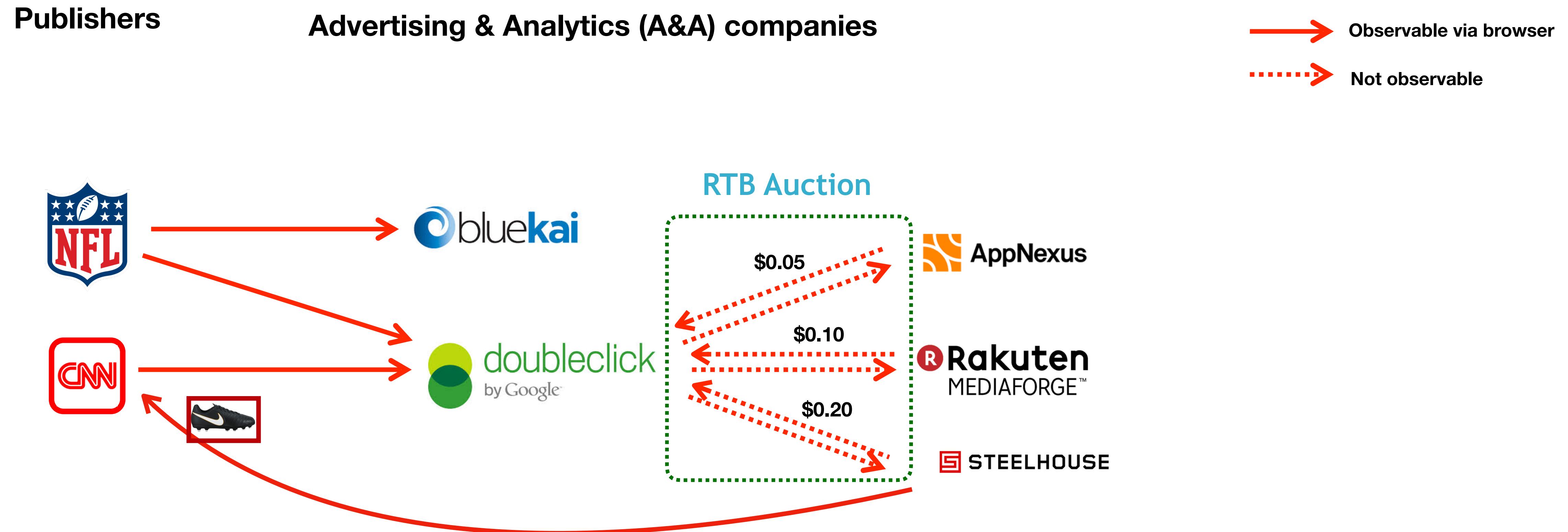
Ecosystem is Moving Toward Real Time Bidding (RTB)



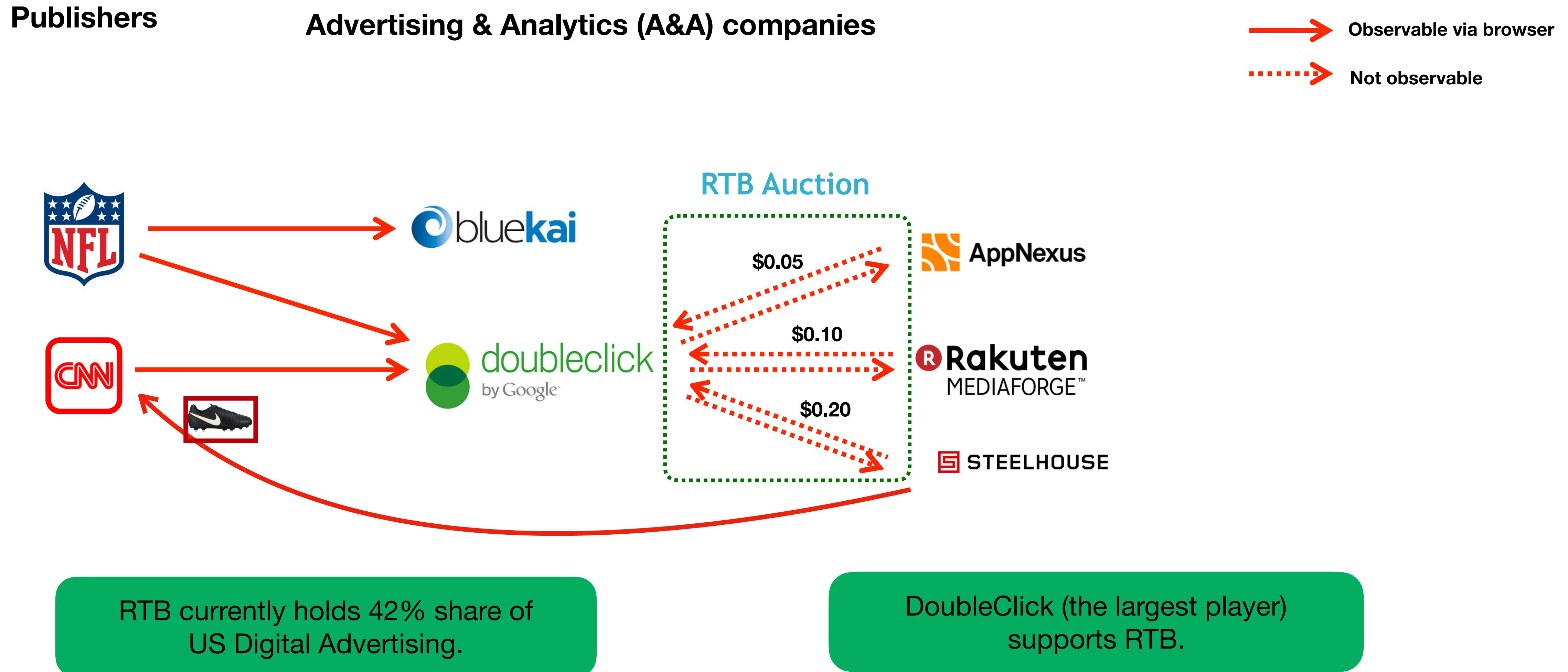
Ecosystem is Moving Toward Real Time Bidding (RTB)



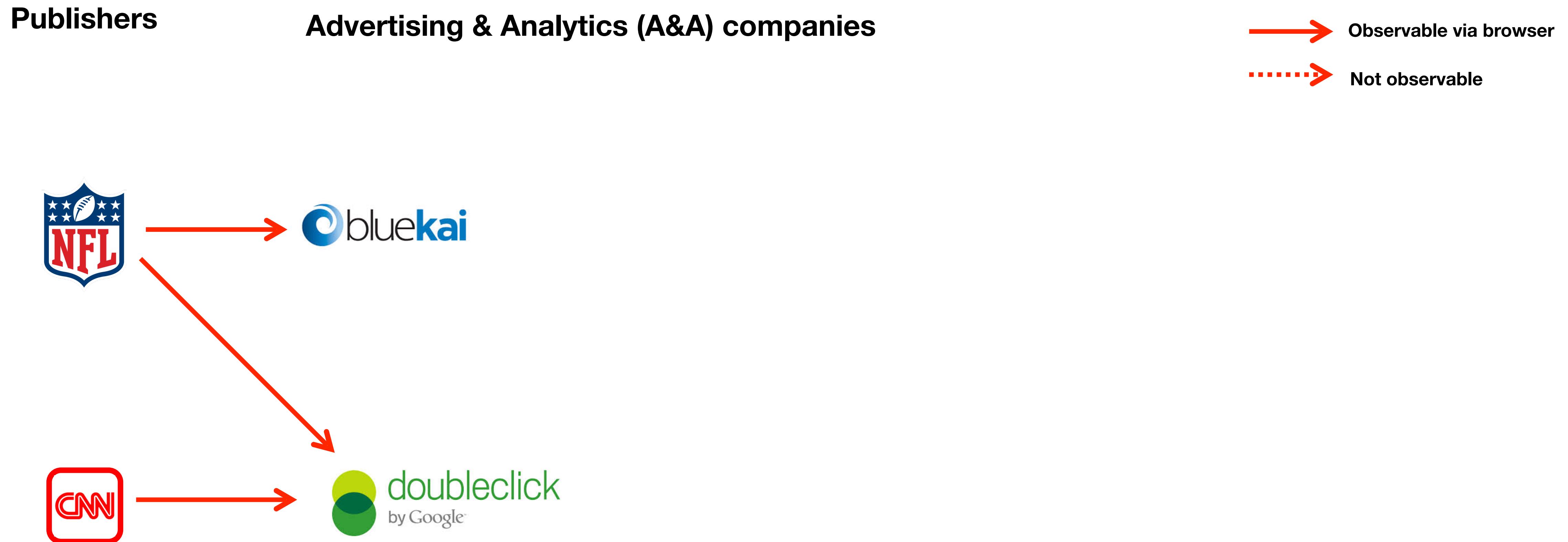
Ecosystem is Moving Toward Real Time Bidding (RTB)



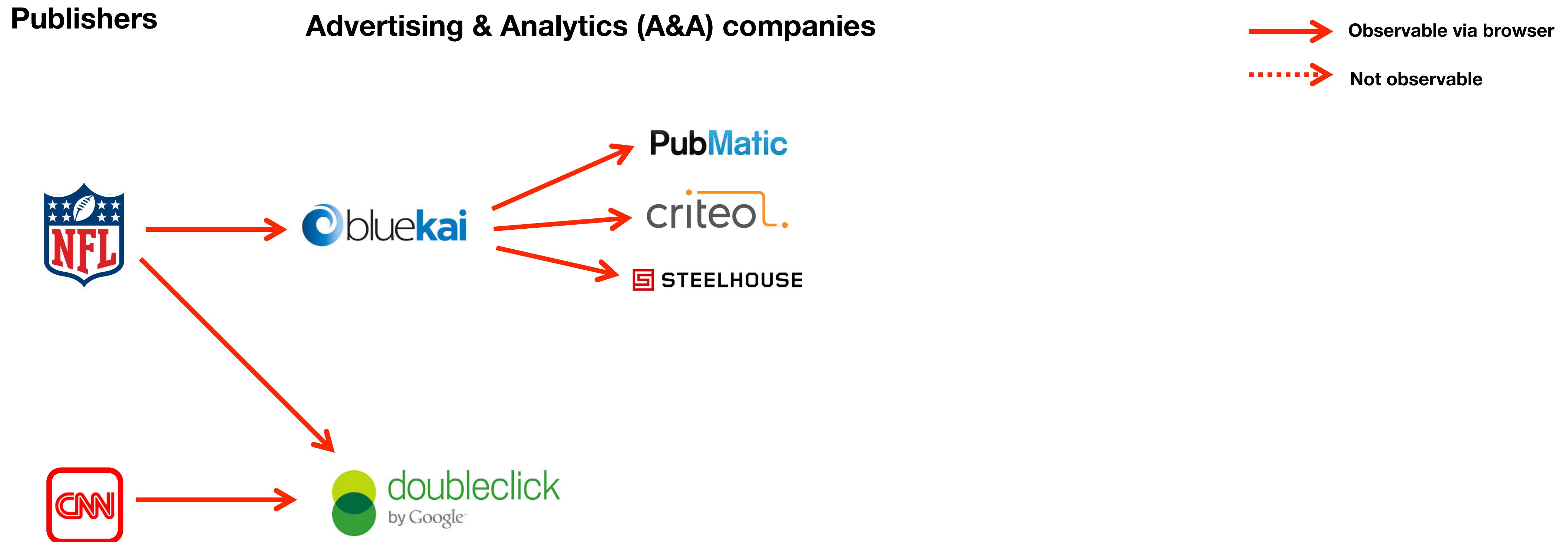
Ecosystem is Moving Toward Real Time Bidding (RTB)



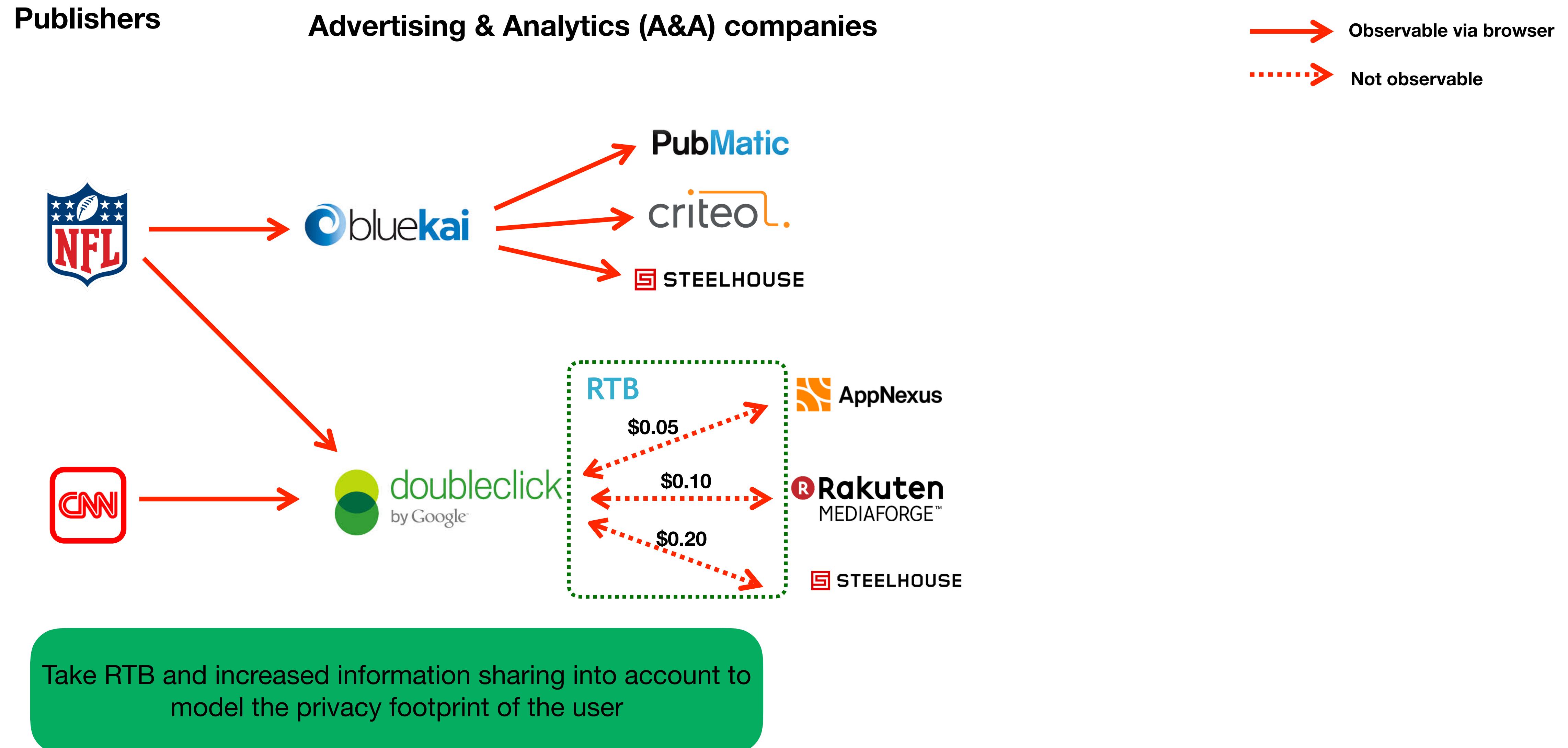
Goal: Model the Privacy Footprint



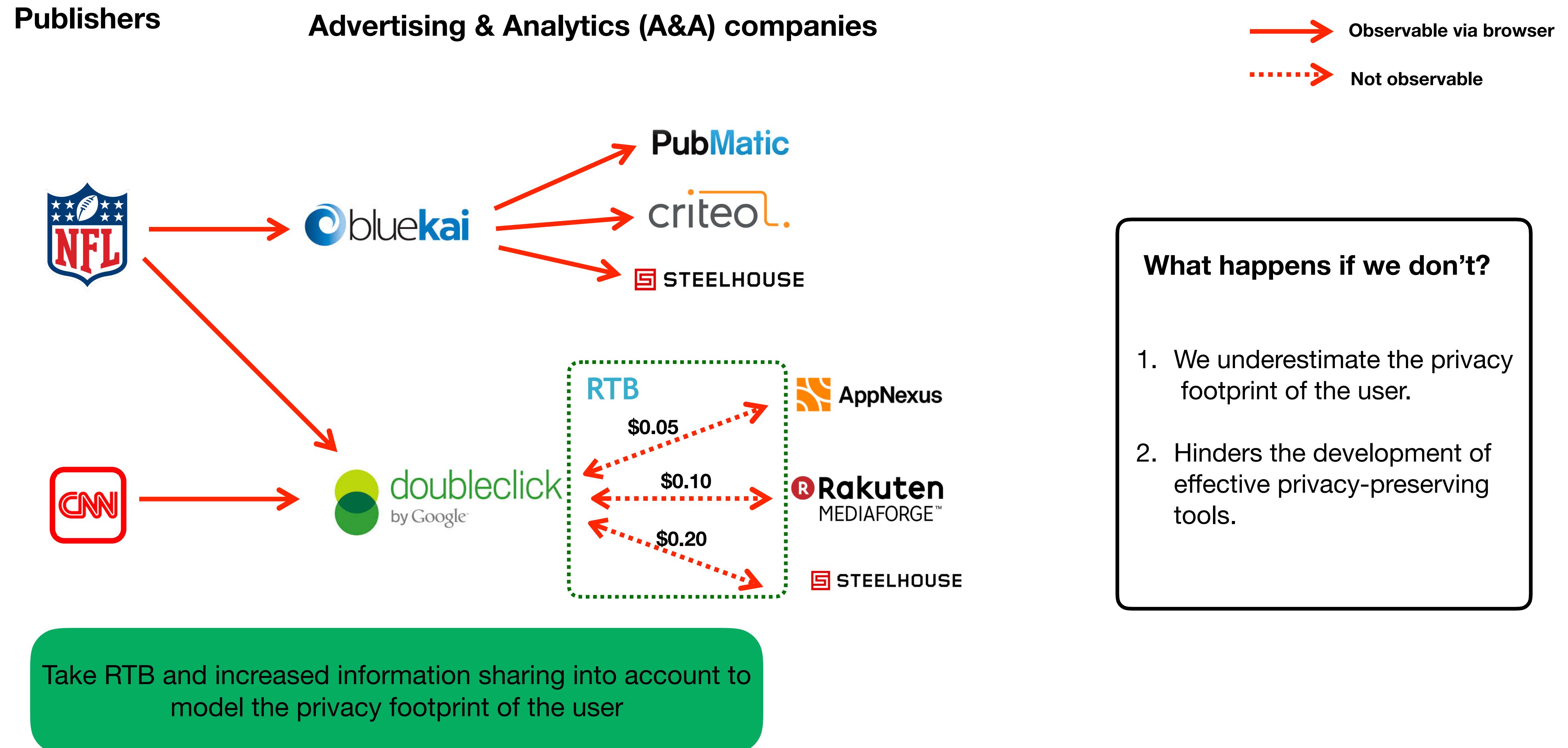
Goal: Model the Privacy Footprint



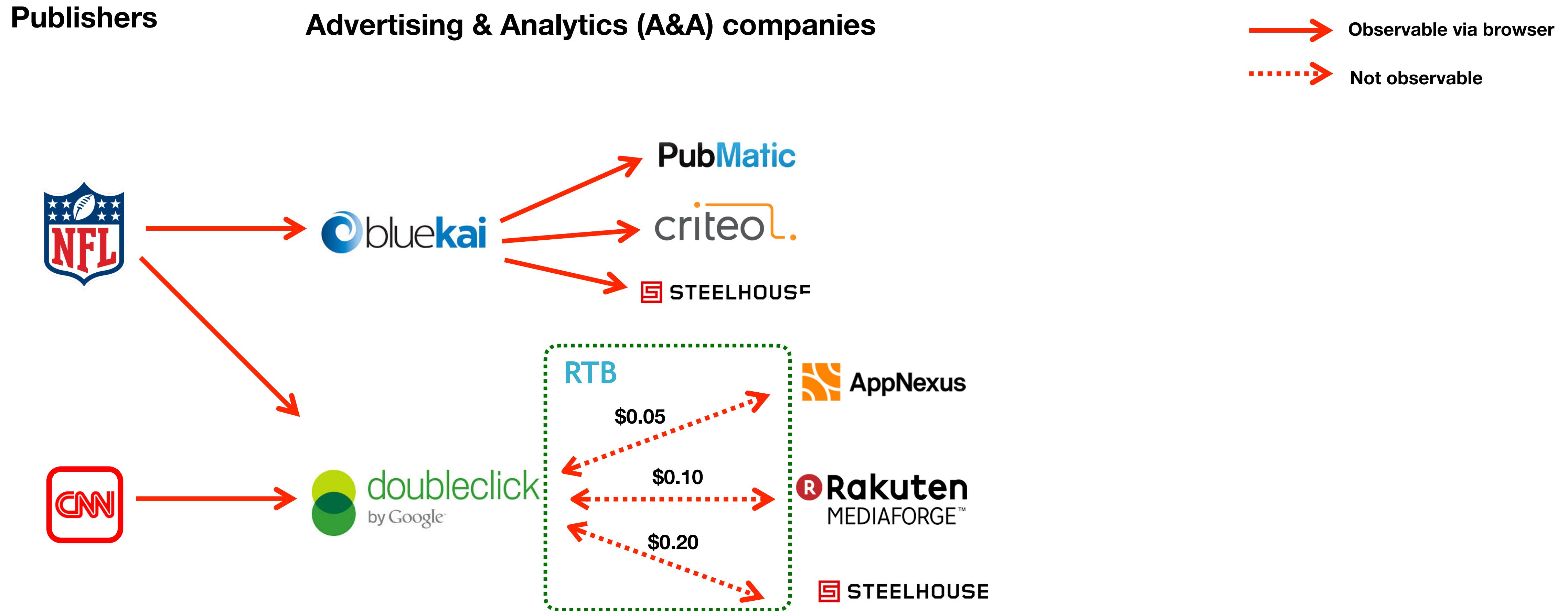
Goal: Model the Privacy Footprint



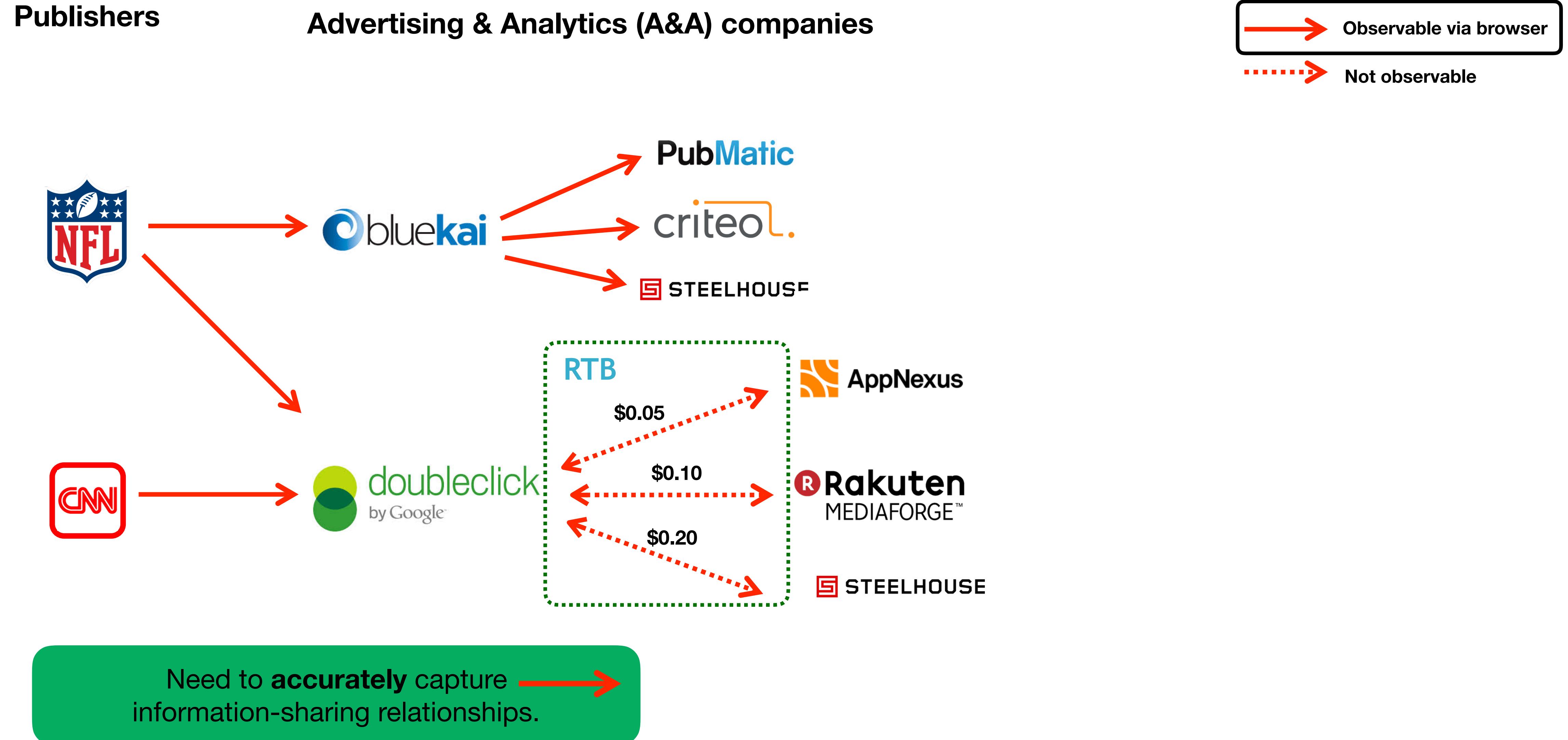
Goal: Model the Privacy Footprint



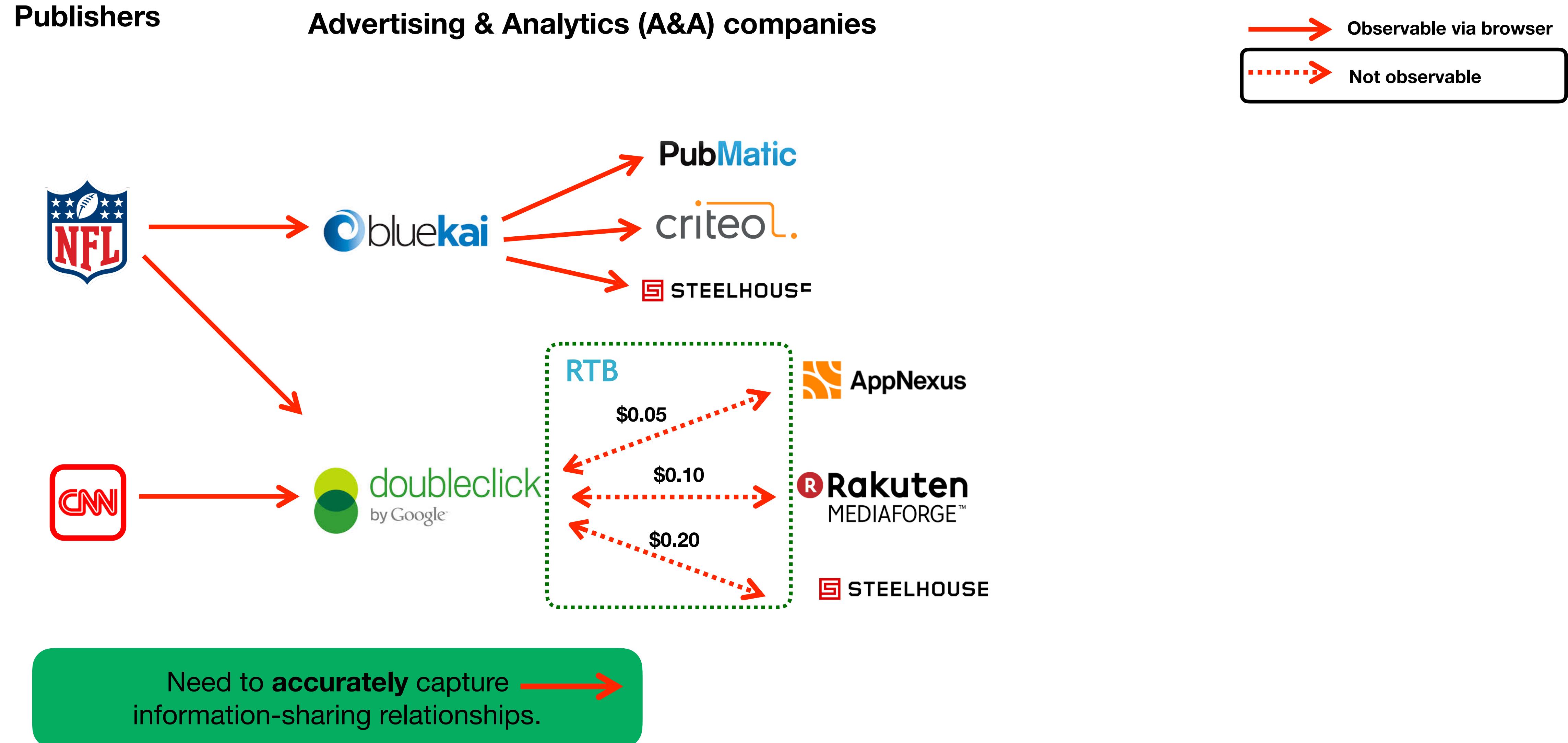
What Do We Need to Model Privacy Footprint?



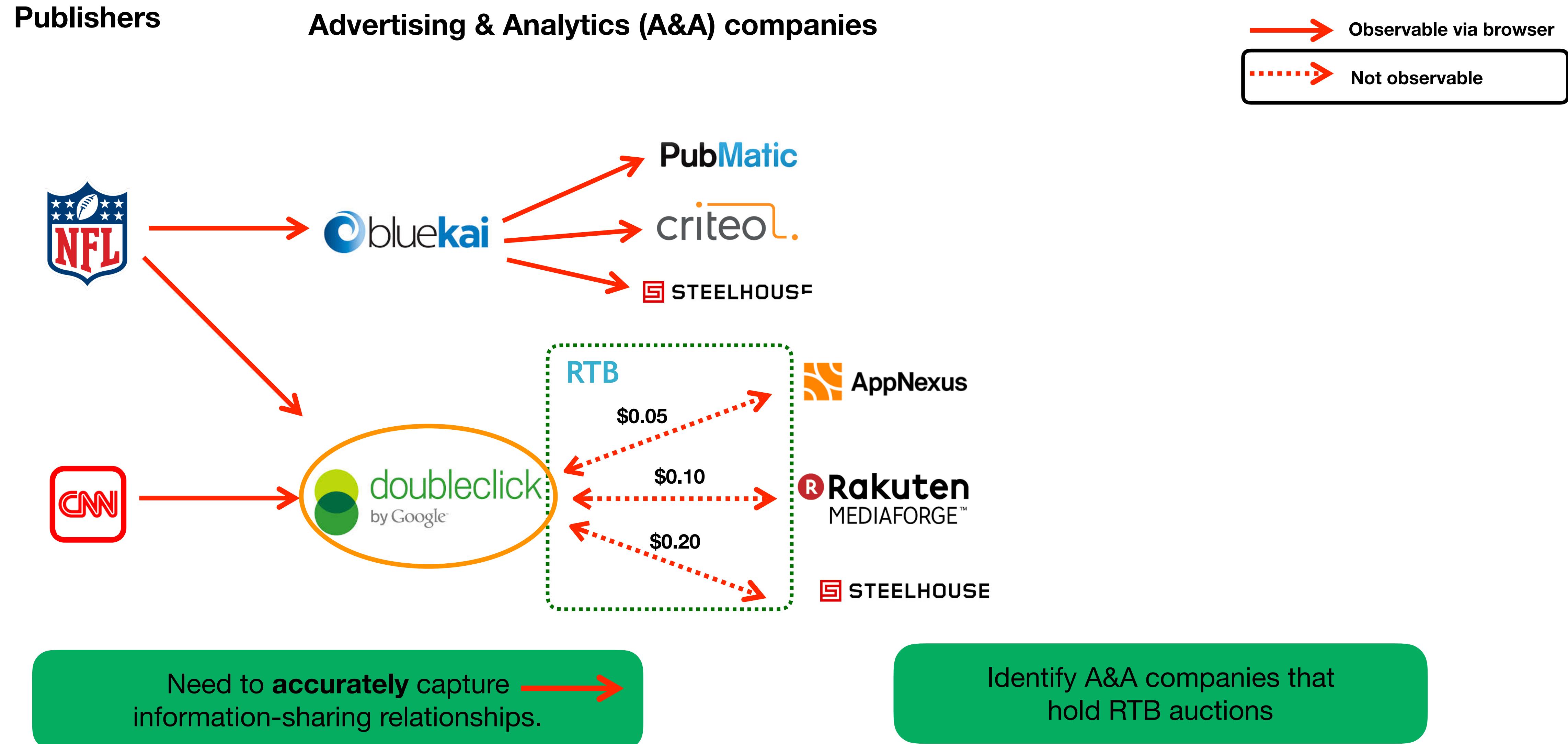
What Do We Need to Model Privacy Footprint?



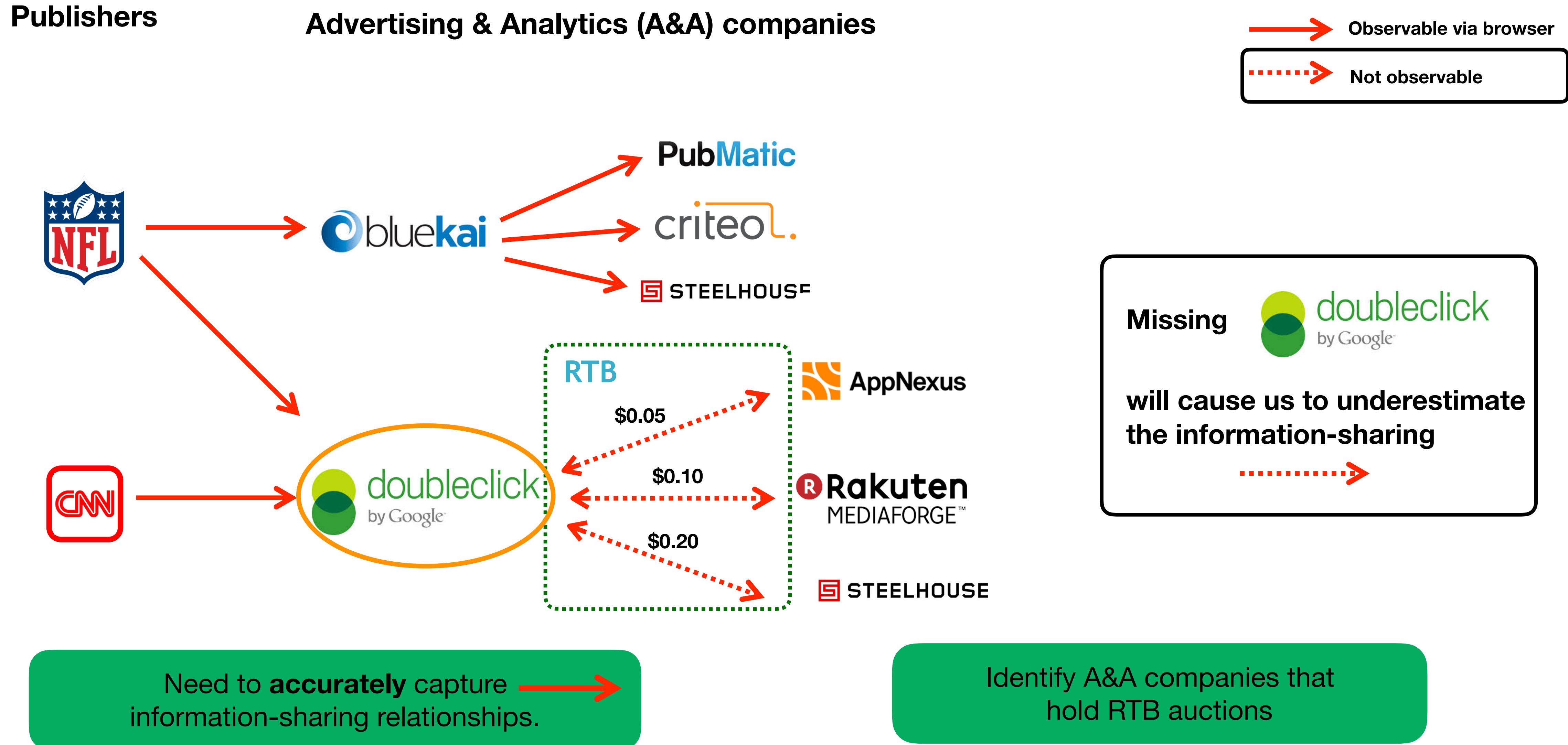
What Do We Need to Model Privacy Footprint?



What Do We Need to Model Privacy Footprint?



What Do We Need to Model Privacy Footprint?



Research on Web Privacy

Research on Web Privacy



Tracking via Cookies

- A&A companies involved
- Prevalence

Research on Web Privacy



Tracking via Cookies

- A&A companies involved
- Prevalence



Fingerprinting

- Browser Extensions
- Local Storage/ Flash LSO
- Canvas Fingerprinting
- Evercookies

Research on Web Privacy



Tracking via Cookies

- A&A companies involved
- Prevalence



Fingerprinting

- Browser Extensions
- Local Storage/ Flash LSO
- Canvas Fingerprinting
- Evercookies



Persona-based Experiments

Research on Web Privacy



Tracking via Cookies

- A&A companies involved
- Prevalence



Fingerprinting

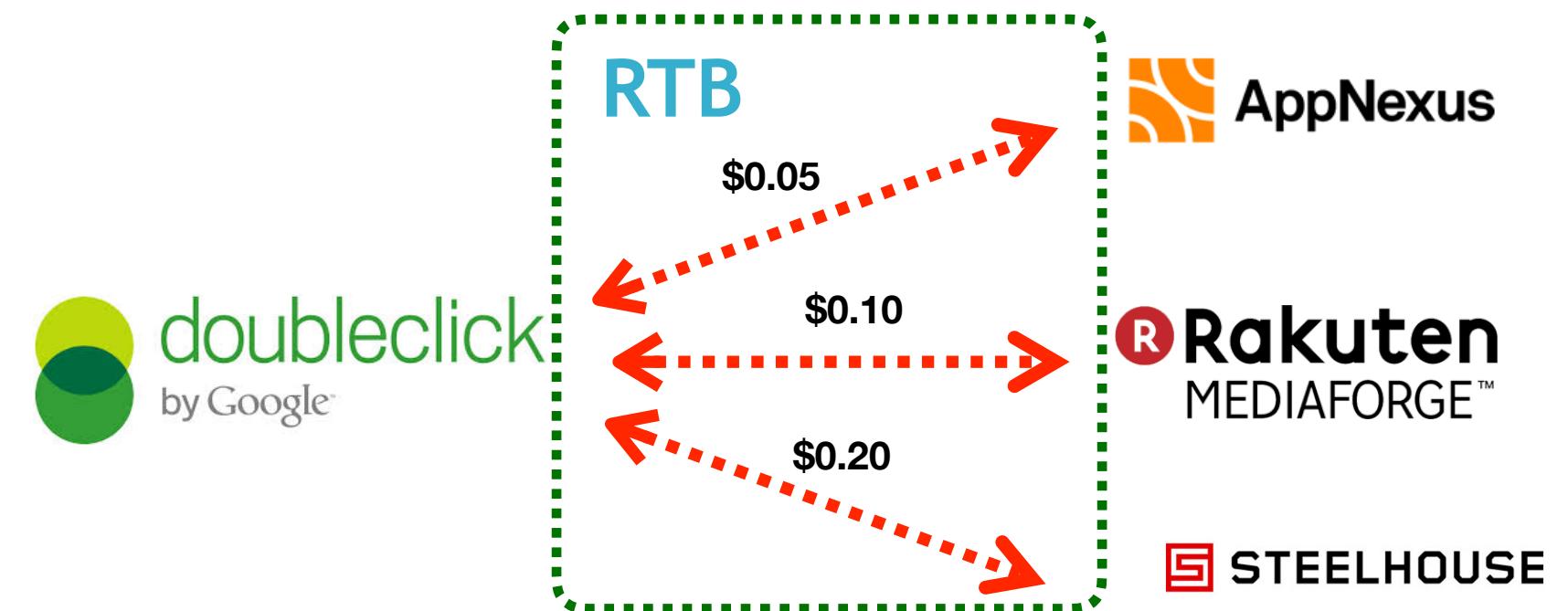
- Browser Extensions
- Local Storage/ Flash LSO
- Canvas Fingerprinting
- Evercookies



Persona-based Experiments



Information sharing with respect to RTB



Goals

To understand privacy implications of RTB for users



Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson [Usenix Security 2016]

A First Look at Content Recommendation Networks

Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson [IMC 2016]



Diffusion of User Tracking Data in the Online Advertising Ecosystem

Muhammad Ahmad Bashir and Christo Wilson [PETS 2018]

How Tracking Companies Circumvented Ad Blockers Using WebSockets

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson [IMC 2018]

Evaluating User Interest Profiles Using Ad Preference Managers

Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar and Christo Wilson [NDSS 2019]



A Longitudinal Analysis of the ads.txt Standard

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson. [IMC 2019]

Key Terms & Definitions

1. **Publishers:** First party websites visited by users (e.g. CNN, BBC, ESPN) 
2. **Advertiser:** Company that want to advertise their products (e.g. Nike, Pepsi) 

Publisher

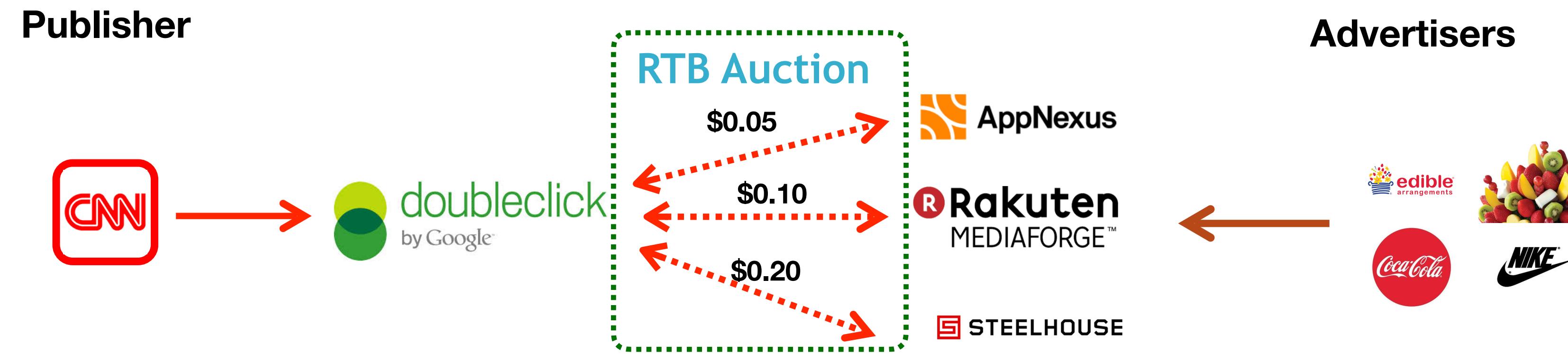


Advertisers



Key Terms & Definitions

1. **Publishers:** First party websites visited by users (e.g. CNN, BBC, ESPN) 
2. **Advertiser:** Company that want to advertise their products (e.g. Nike, Pepsi) 



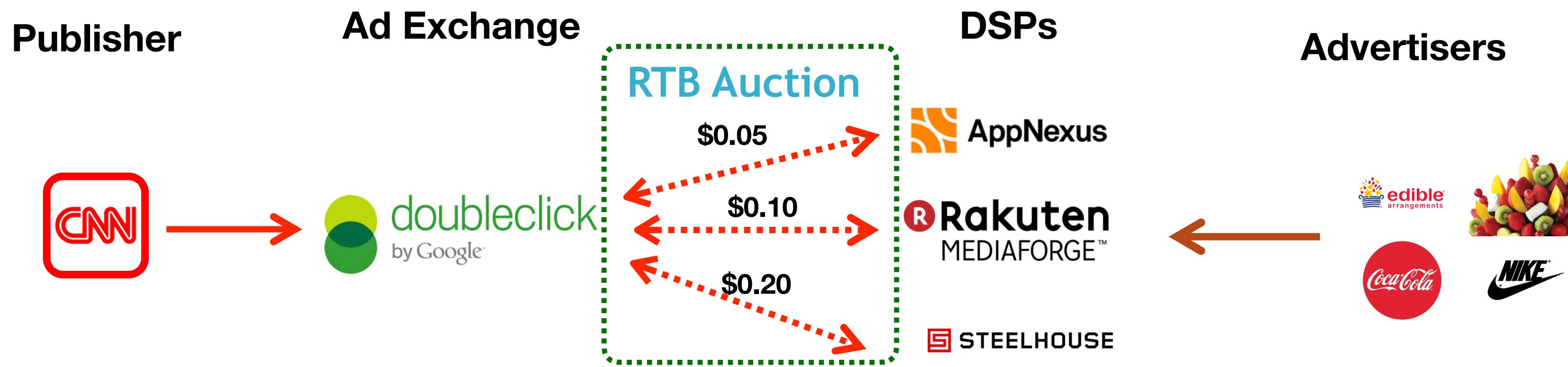
Key Terms & Definitions

1. **Publisher:** First party websites visited by users (e.g. CNN, BBC, ESPN) 
2. **Advertiser:** Company that want to advertise their products (e.g. Nike, Pepsi) 
3. **Ad Exchange:** A&A company that holds the auction and solicits bids 
4. **Demand Side Platform (DSP):** A&A company that bids on advertisers' behalf 



Key Terms & Definitions

1. **Publishers:** First party websites visited by users (e.g. CNN, BBC, ESPN) 
2. **Advertiser:** Company that want to advertise their products (e.g. Nike, Pepsi) 
3. **Ad Exchange:** A&A company that holds the auction and solicits bids 
4. **Demand Side Platform (DSP):** A&A company that bids on advertisers' behalf 
5. **A&A:** All Advertising and Analytics related companies / domains (e.g. Google, Criteo)
6. **Impressions:** Page visits users generate on websites



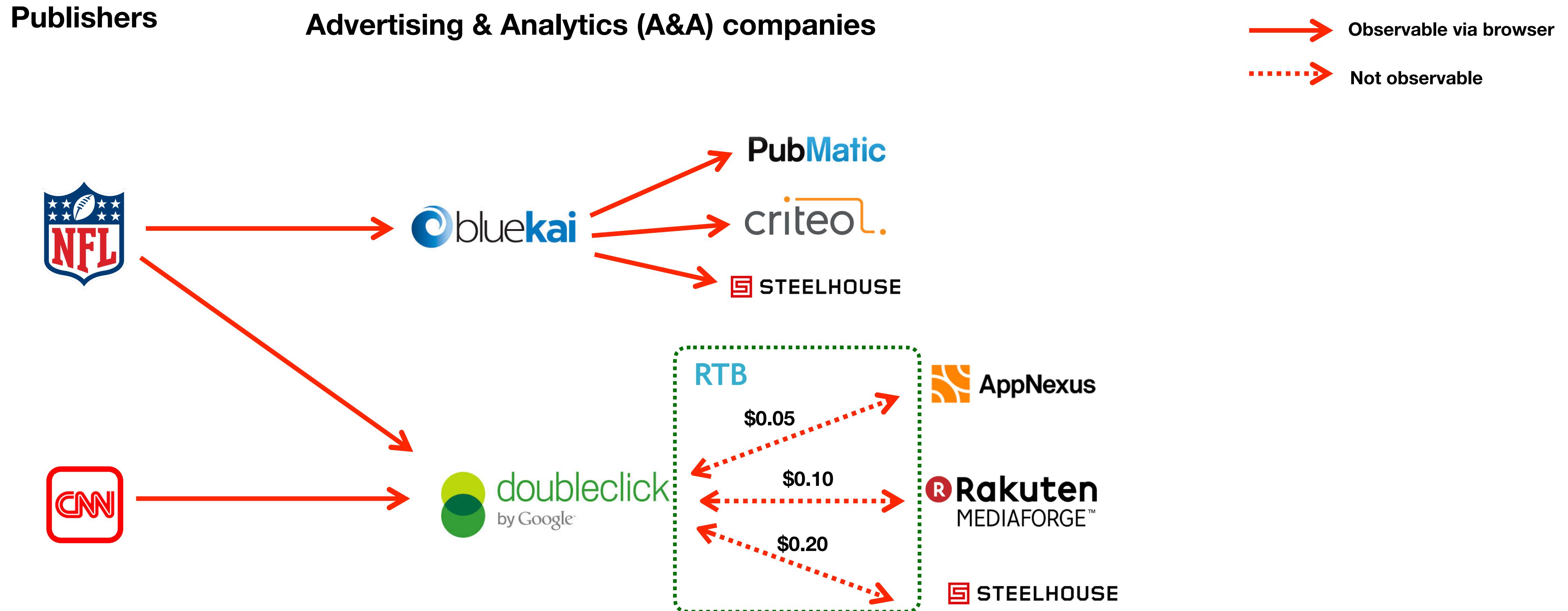
Thesis Outline

Detecting Information Sharing

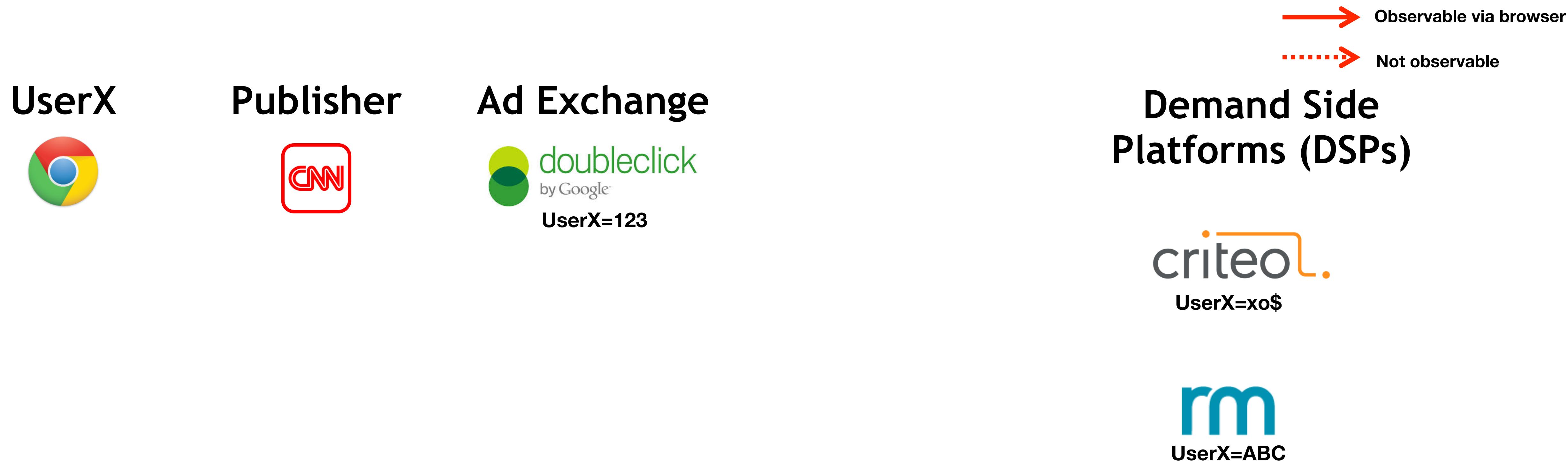
Isolating List of Ad Exchanges

Modeling User's Privacy Digital Footprint

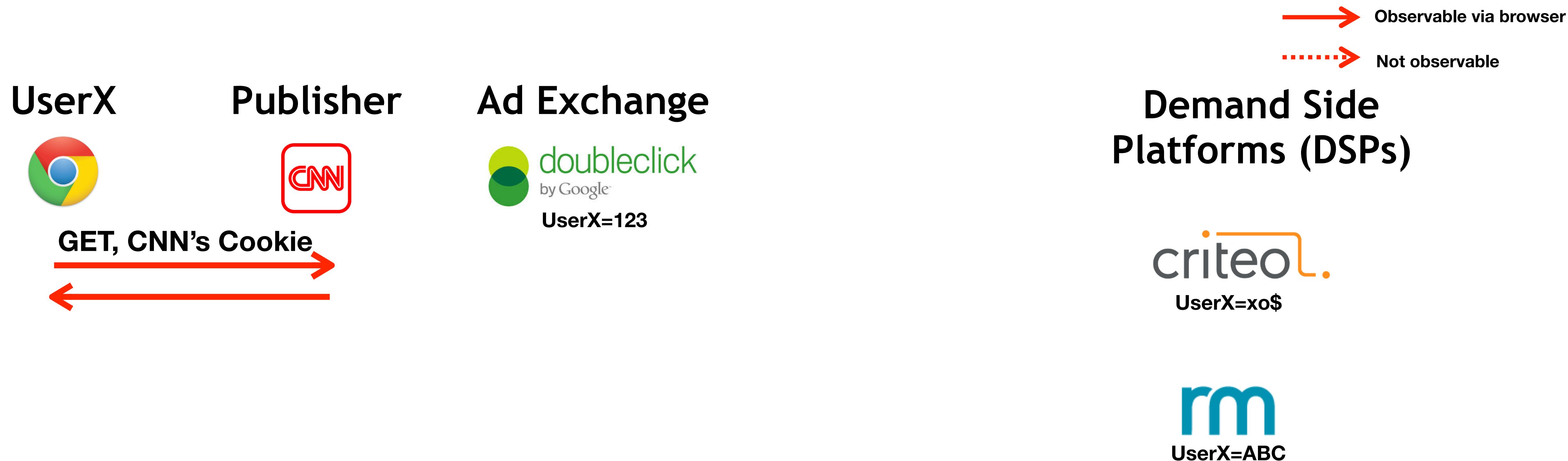
RTB Has Fueled Information Sharing



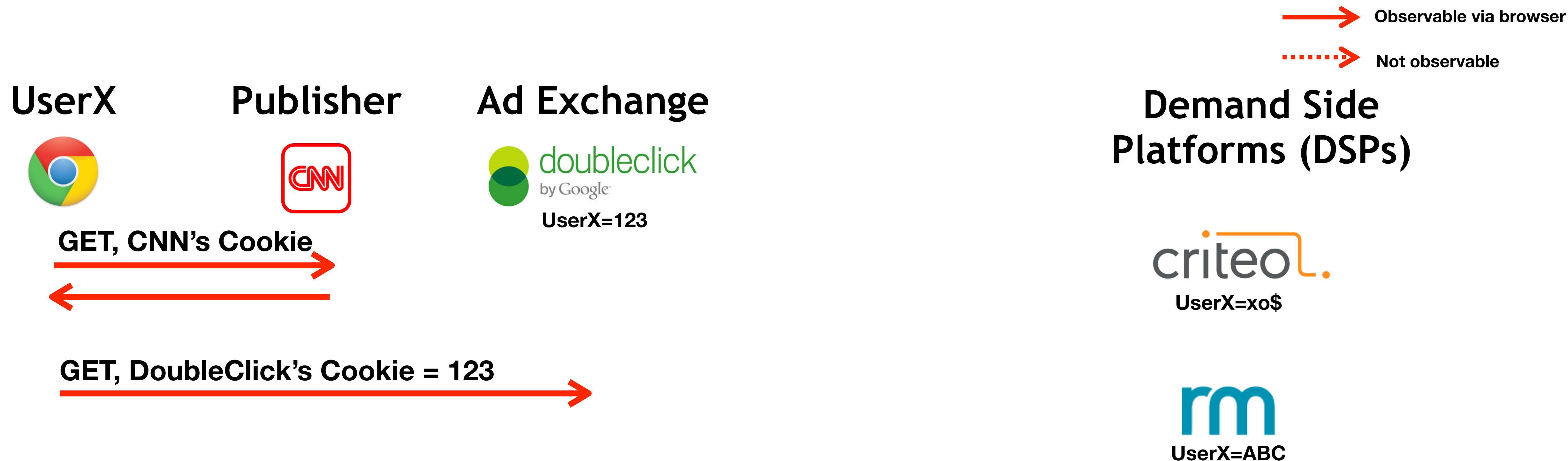
An Example of Real Time Bidding (RTB) Auction



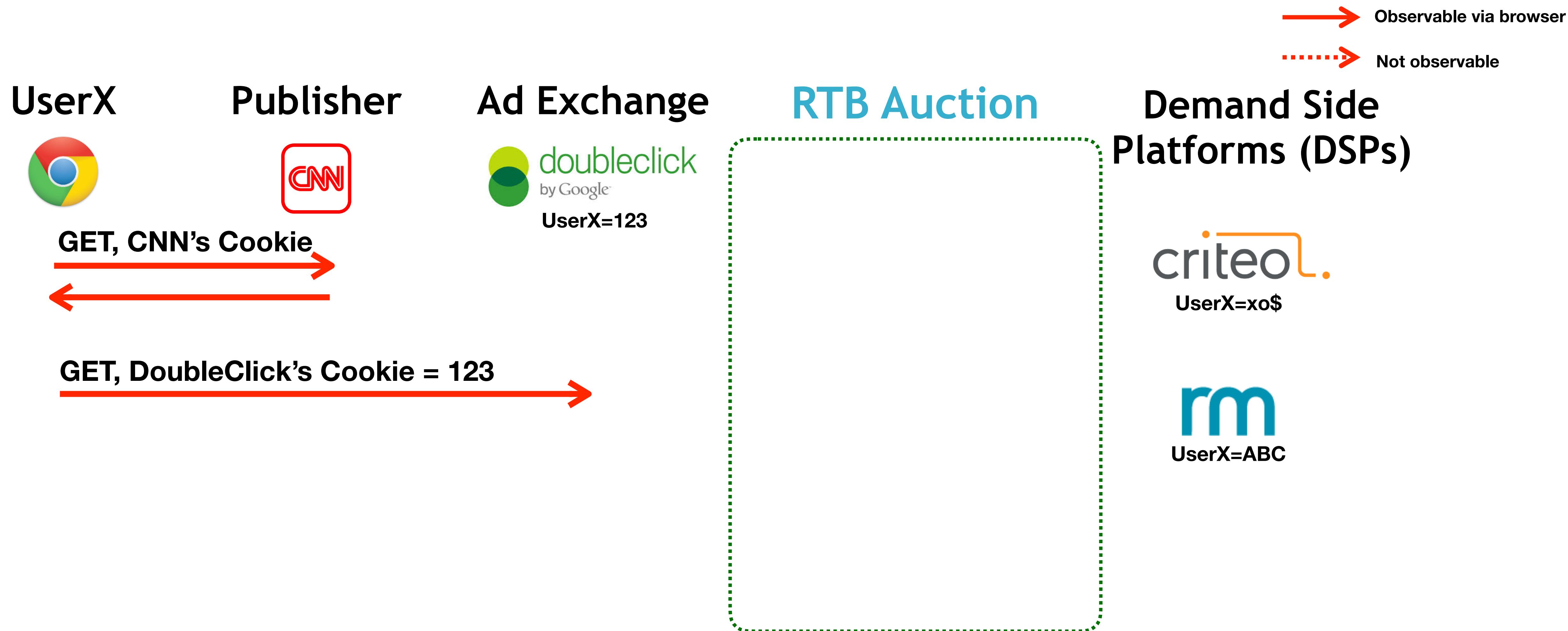
An Example of Real Time Bidding (RTB) Auction



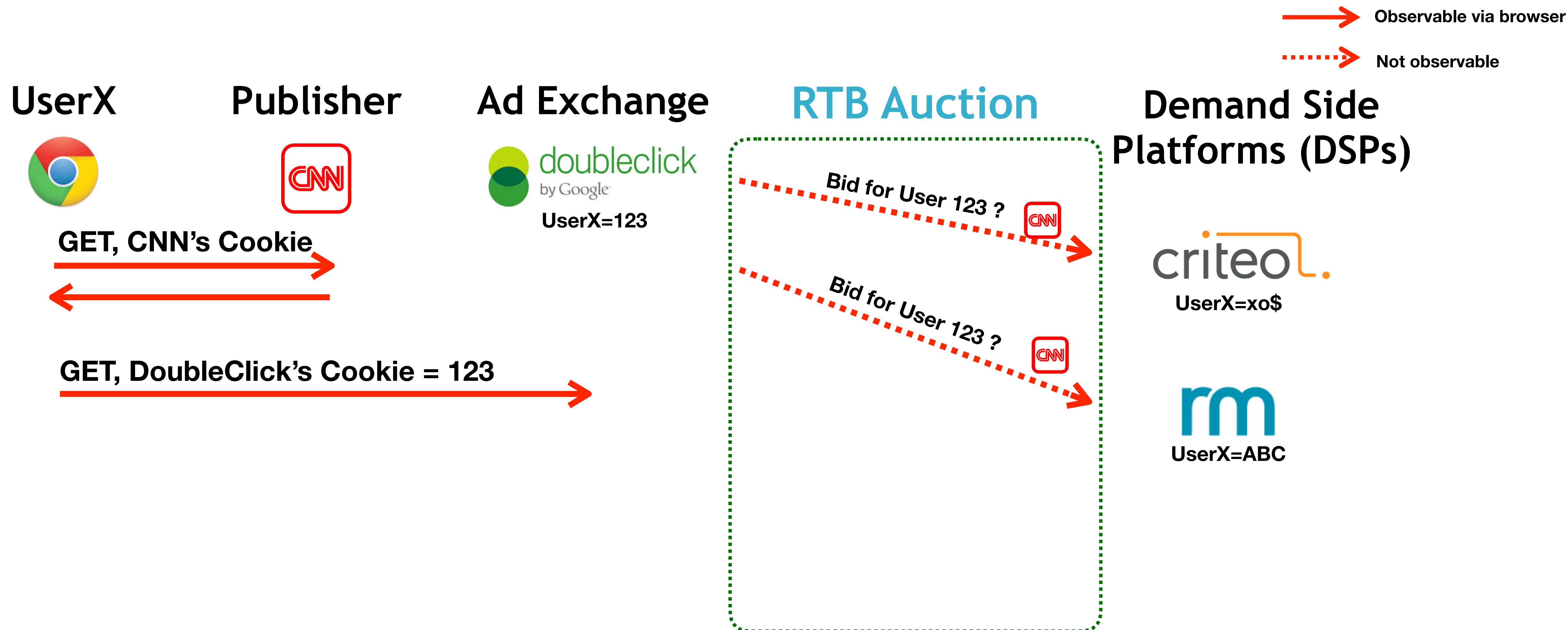
An Example of Real Time Bidding (RTB) Auction



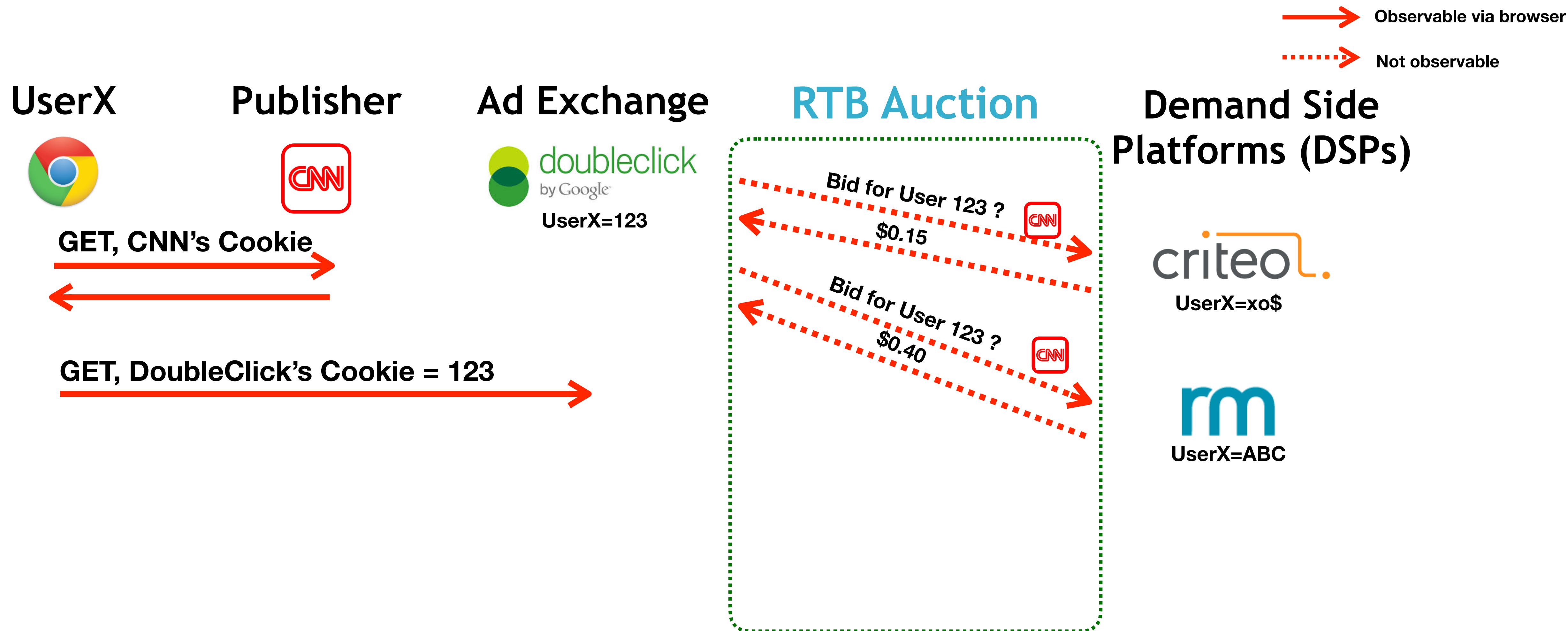
An Example of Real Time Bidding (RTB) Auction



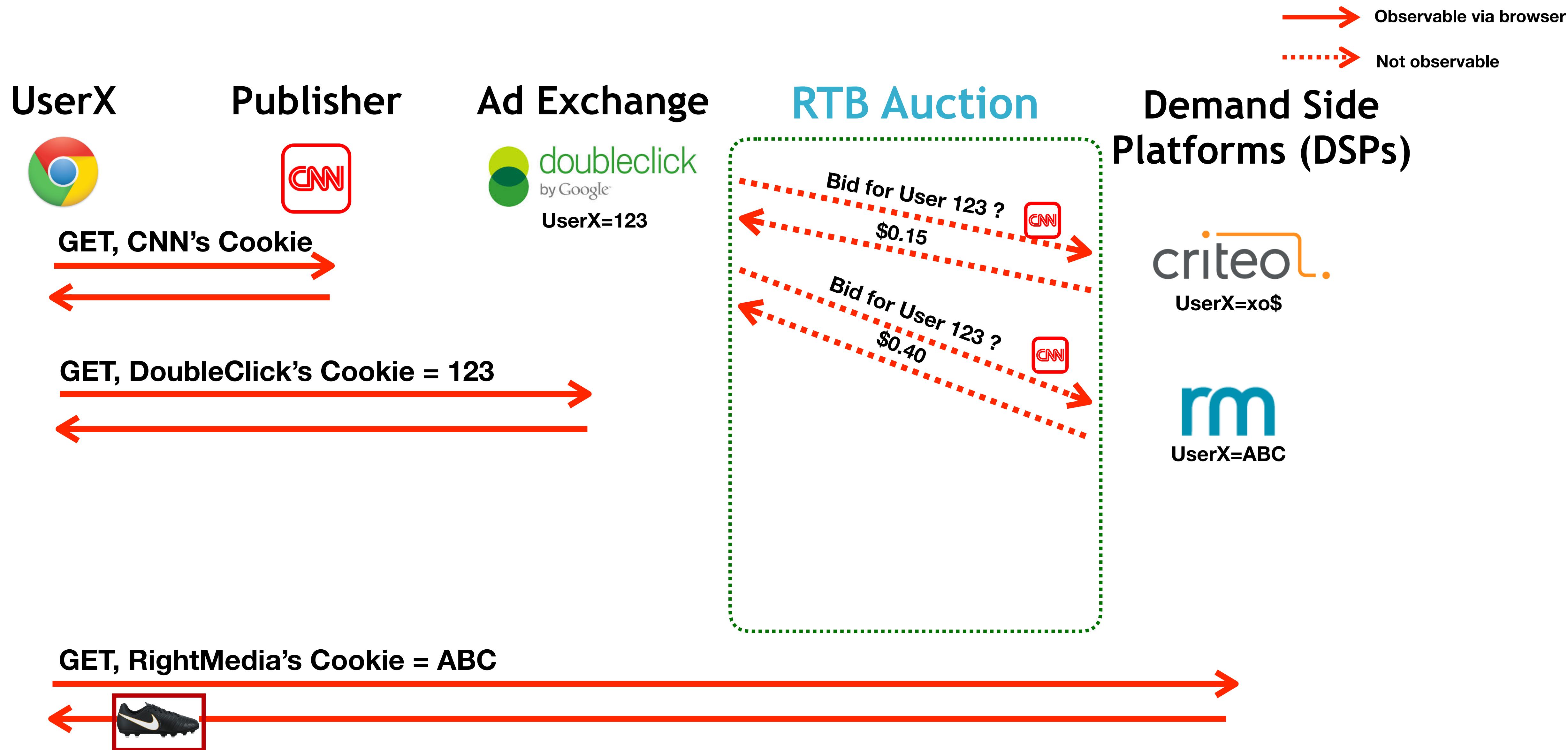
An Example of Real Time Bidding (RTB) Auction



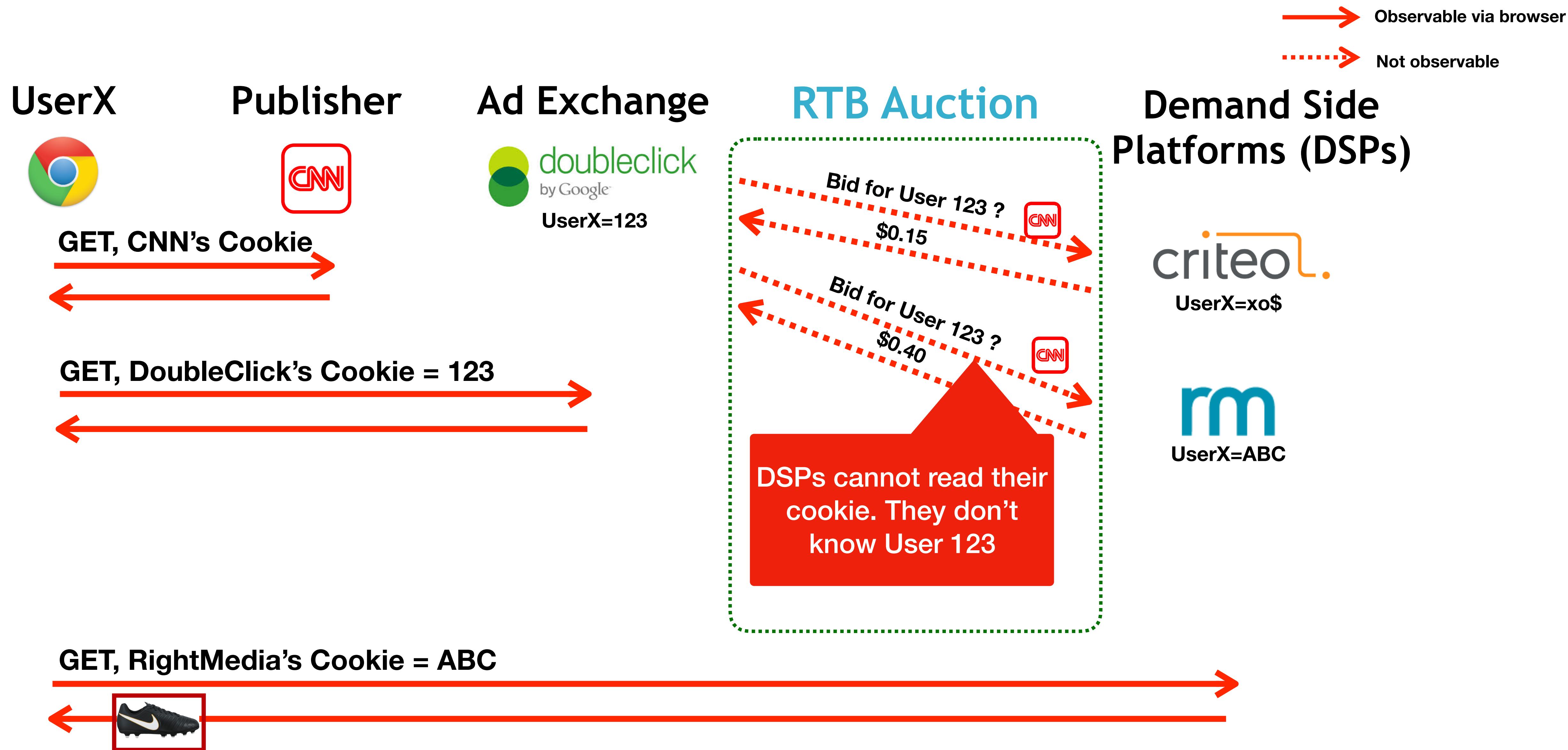
An Example of Real Time Bidding (RTB) Auction



An Example of Real Time Bidding (RTB) Auction



An Example of Real Time Bidding (RTB) Auction



Cookie Matching

Key problem: DSPs (bidders) cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains

Cookie Matching

Key problem: DSPs (bidders) cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



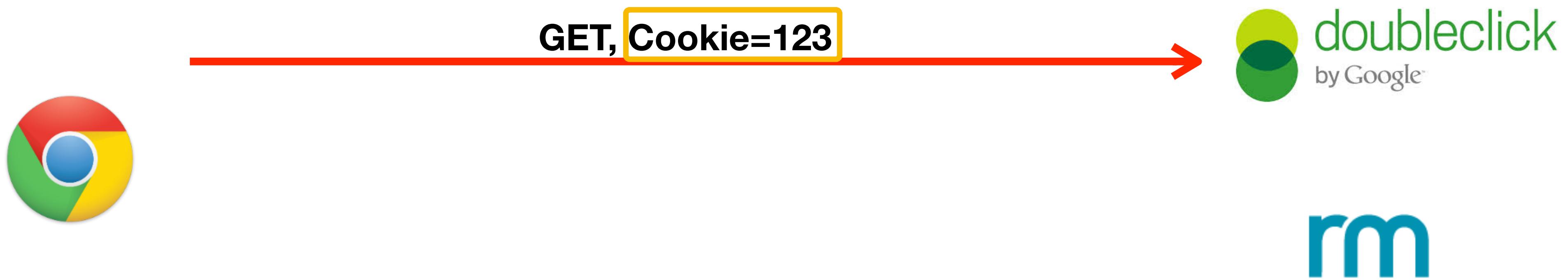
Cookie Matching

Key problem: DSPs (bidders) cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



Cookie Matching

- Key problem:** DSPs (bidders) cannot read their cookies in the RTB auction
- How can they submit reasonable bids if they cannot identify the user?

Solution: cookie matching

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



Cookie Matching

Key problem: DSPs (bidders) cannot read their cookies in the RTB auction

- How can they submit reasonable bids if they cannot identify the user?

Solution: [cookie matching](#)

- Also known as cookie synching
- Process of linking the identifiers used by two ad domains

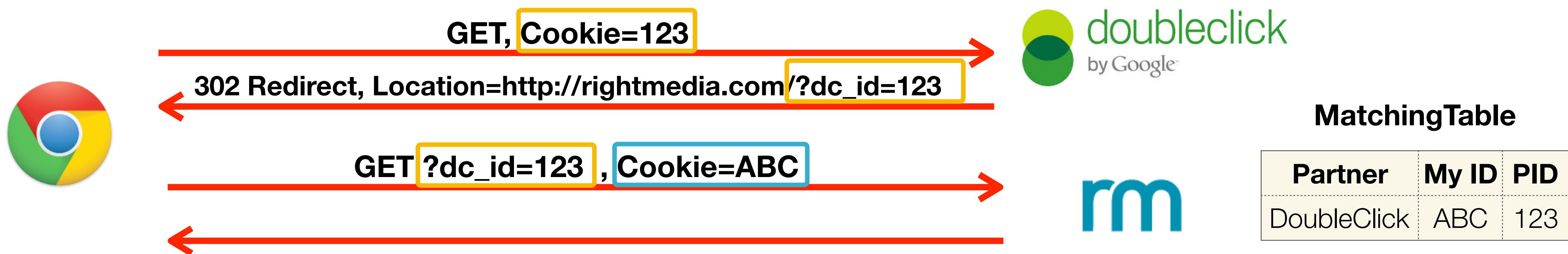


Cookie Matching

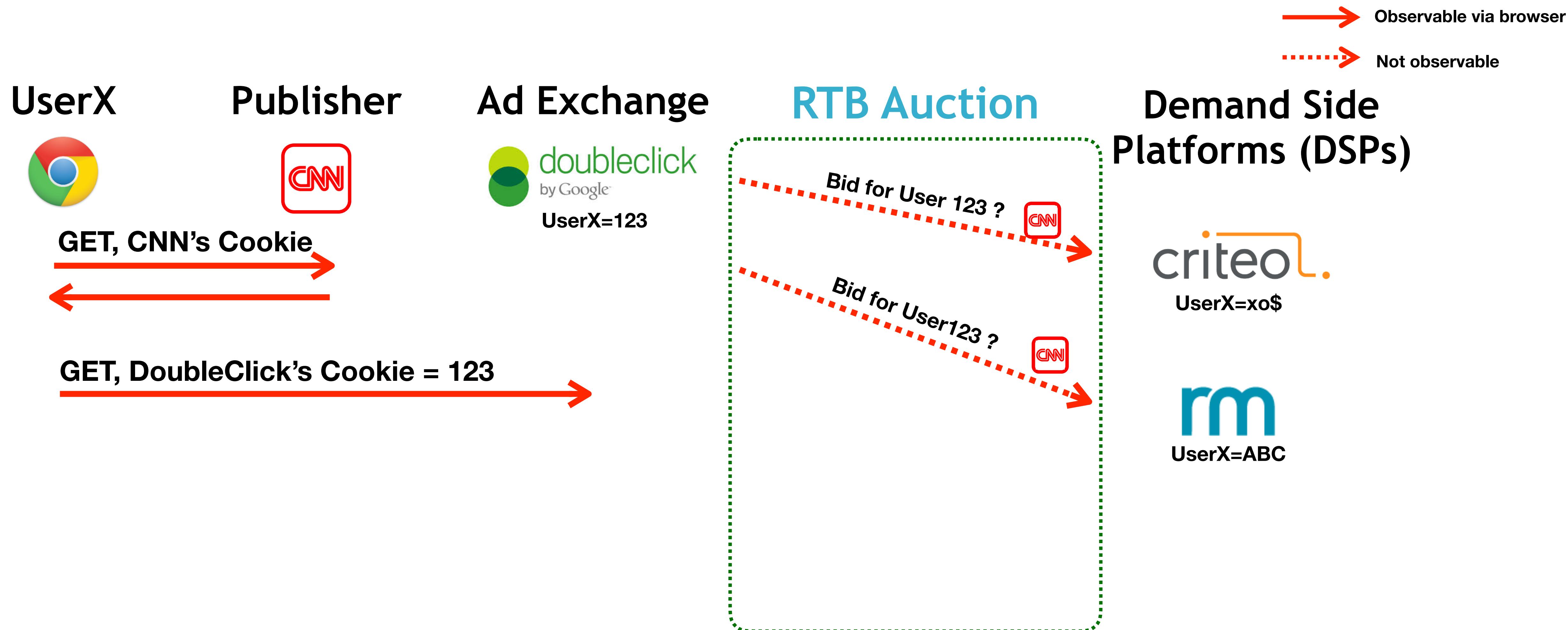
- Key problem:** DSPs (bidders) cannot read their cookies in the RTB auction
- How can they submit reasonable bids if they cannot identify the user?

Solution: cookie matching

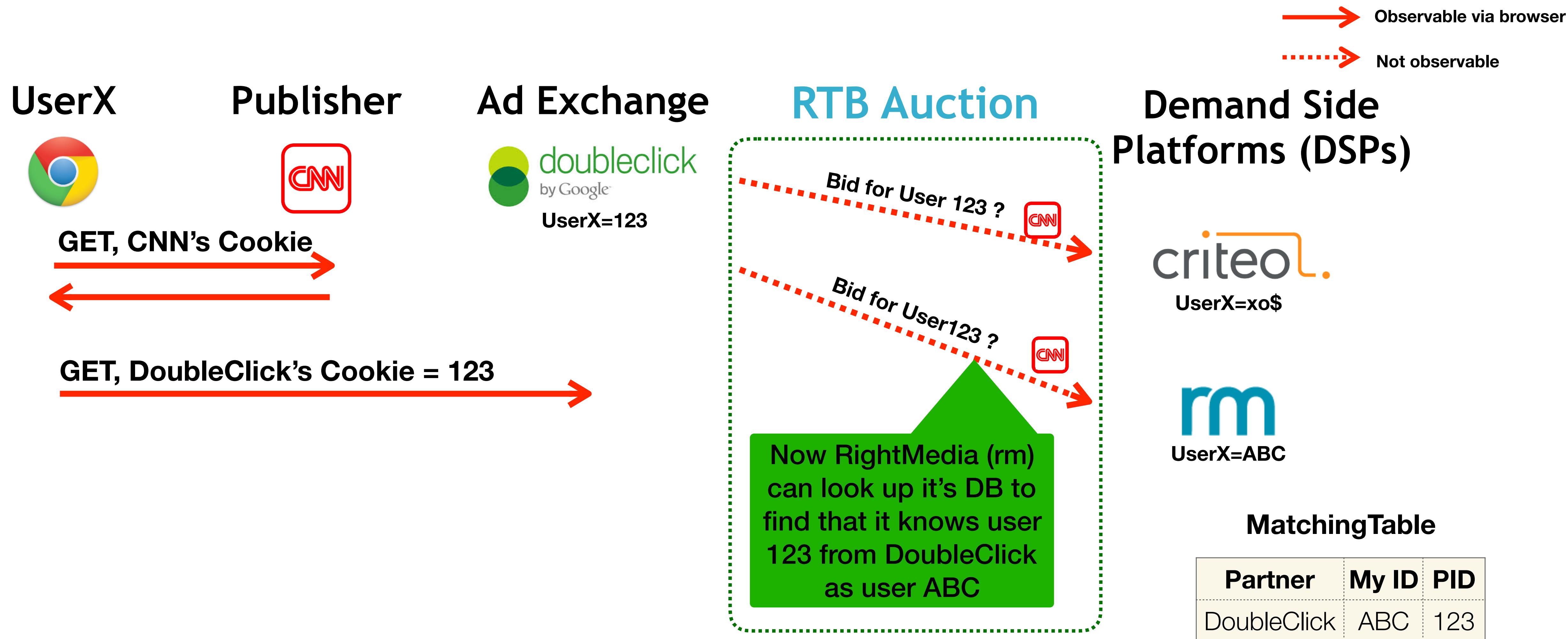
- Also known as cookie synching
- Process of linking the identifiers used by two ad domains



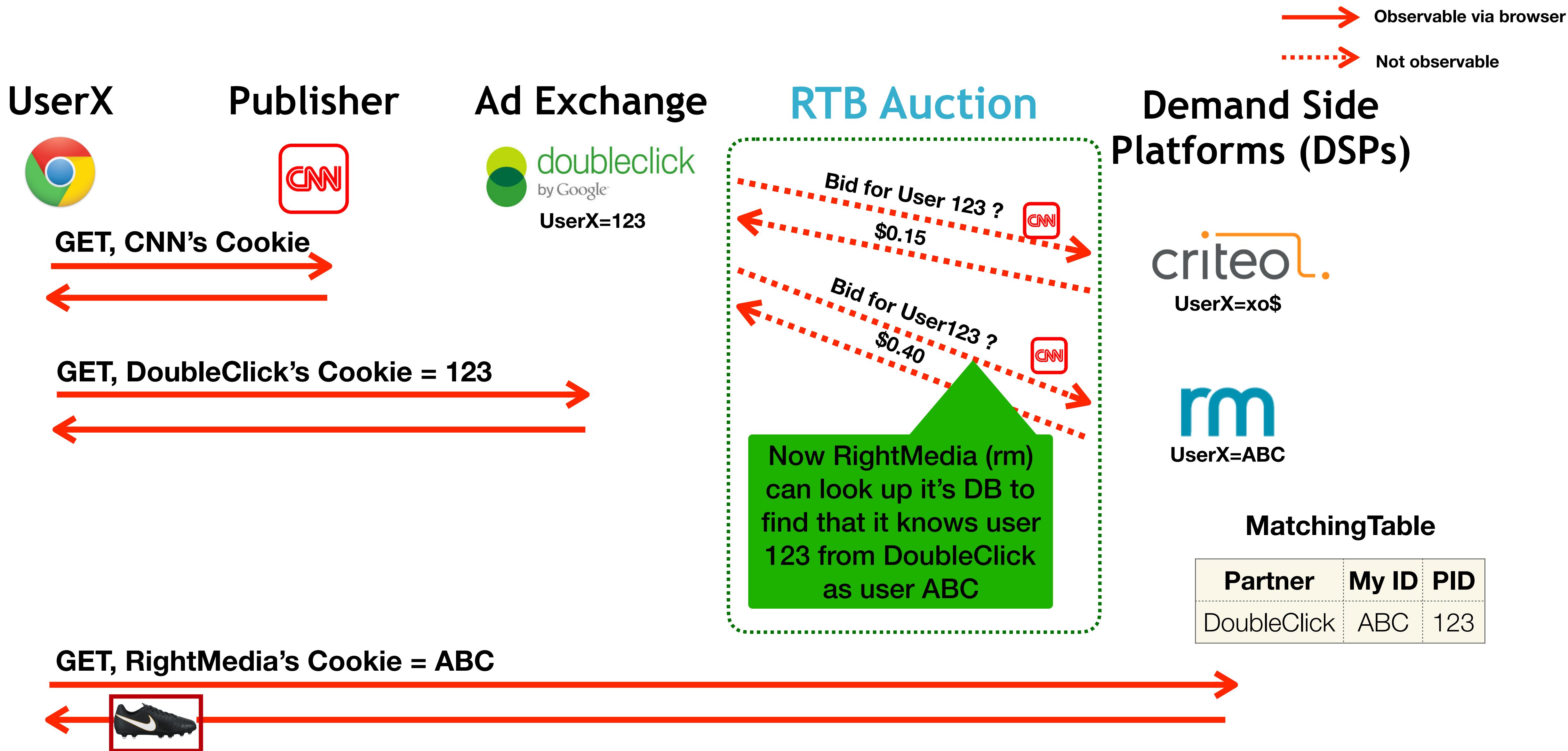
An Example of Real Time Bidding (RTB) Auction



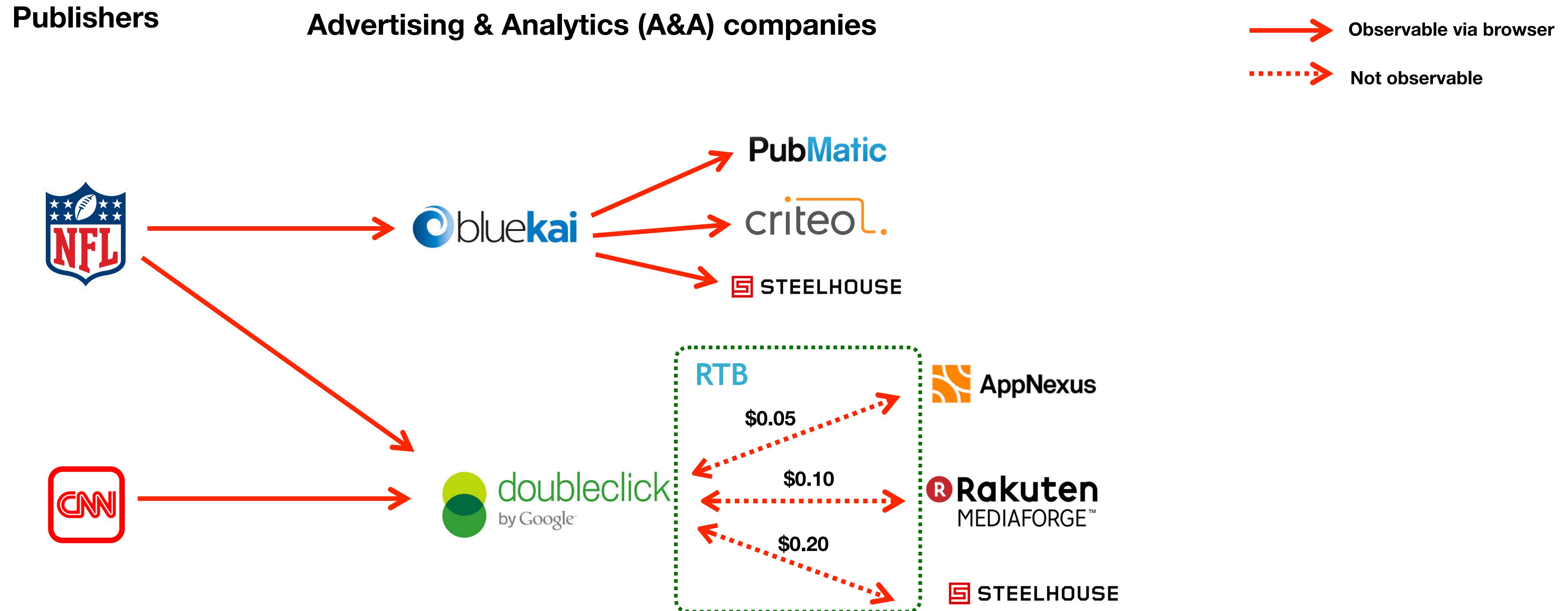
An Example of Real Time Bidding (RTB) Auction



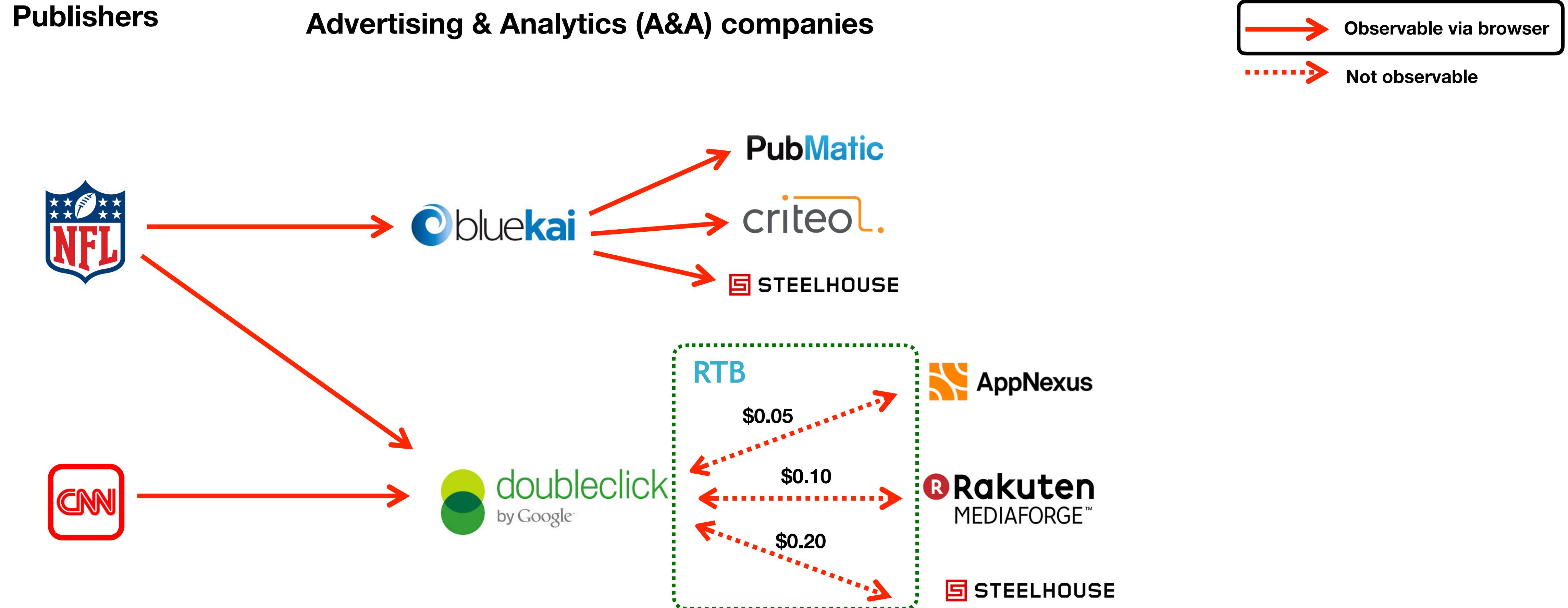
An Example of Real Time Bidding (RTB) Auction



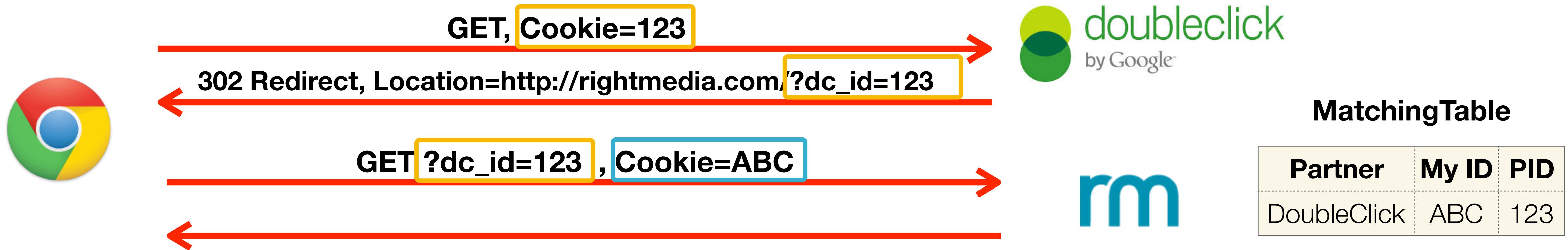
Need To Capture Information Sharing Relationships



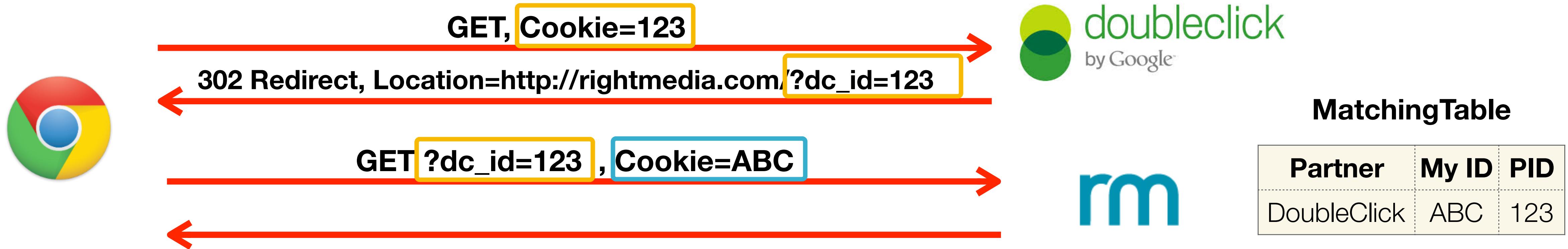
Need To Capture Information Sharing Relationships



Using Redirects to Determine Information Sharing?



Using Redirects to Determine Information Sharing?



Can't we just rely on HTTP headers to look for redirects?

- Acar *et al.* found hundreds of domains passing identifiers to each other
- Olejnik *et al.* found 125 exchanges matching cookies
- Falahrastegar *et al.* analyzed clusters of exchanges that share the exact same cookies

It's a decent approach — but not without limitations

Limitation #1: Server Side Matching

Limitation #1: Server Side Matching

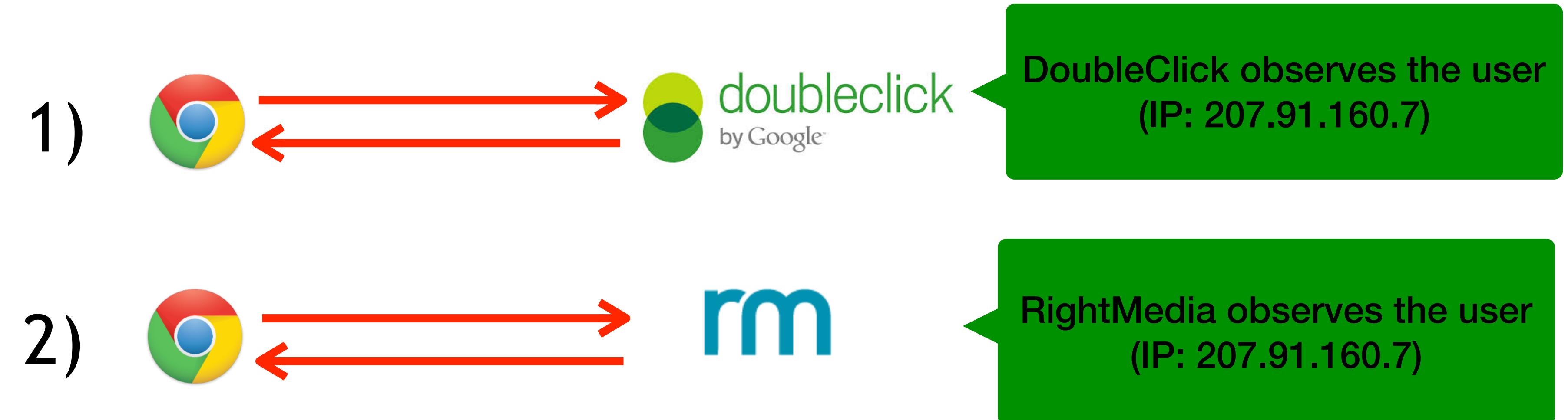
1)



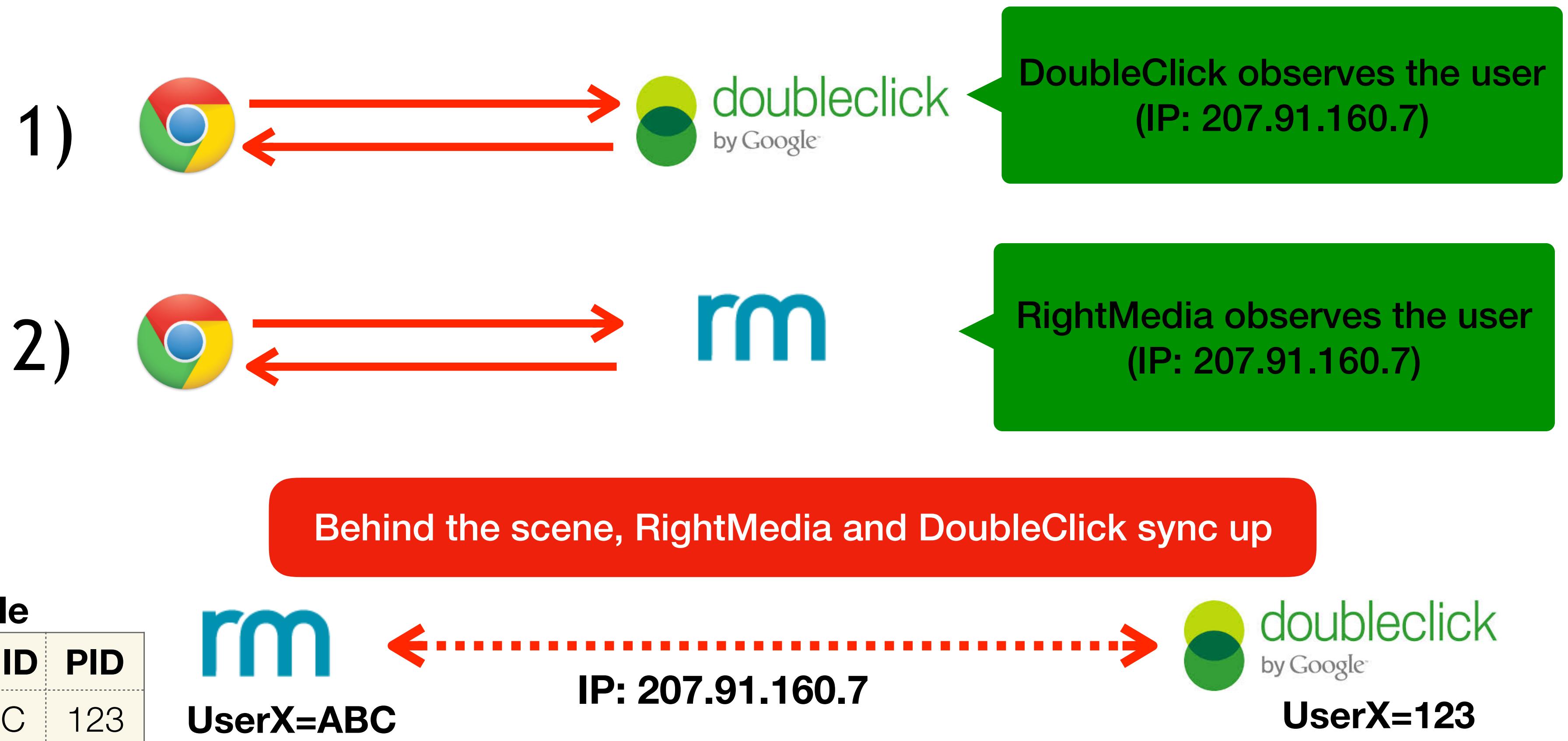
Limitation #1: Server Side Matching



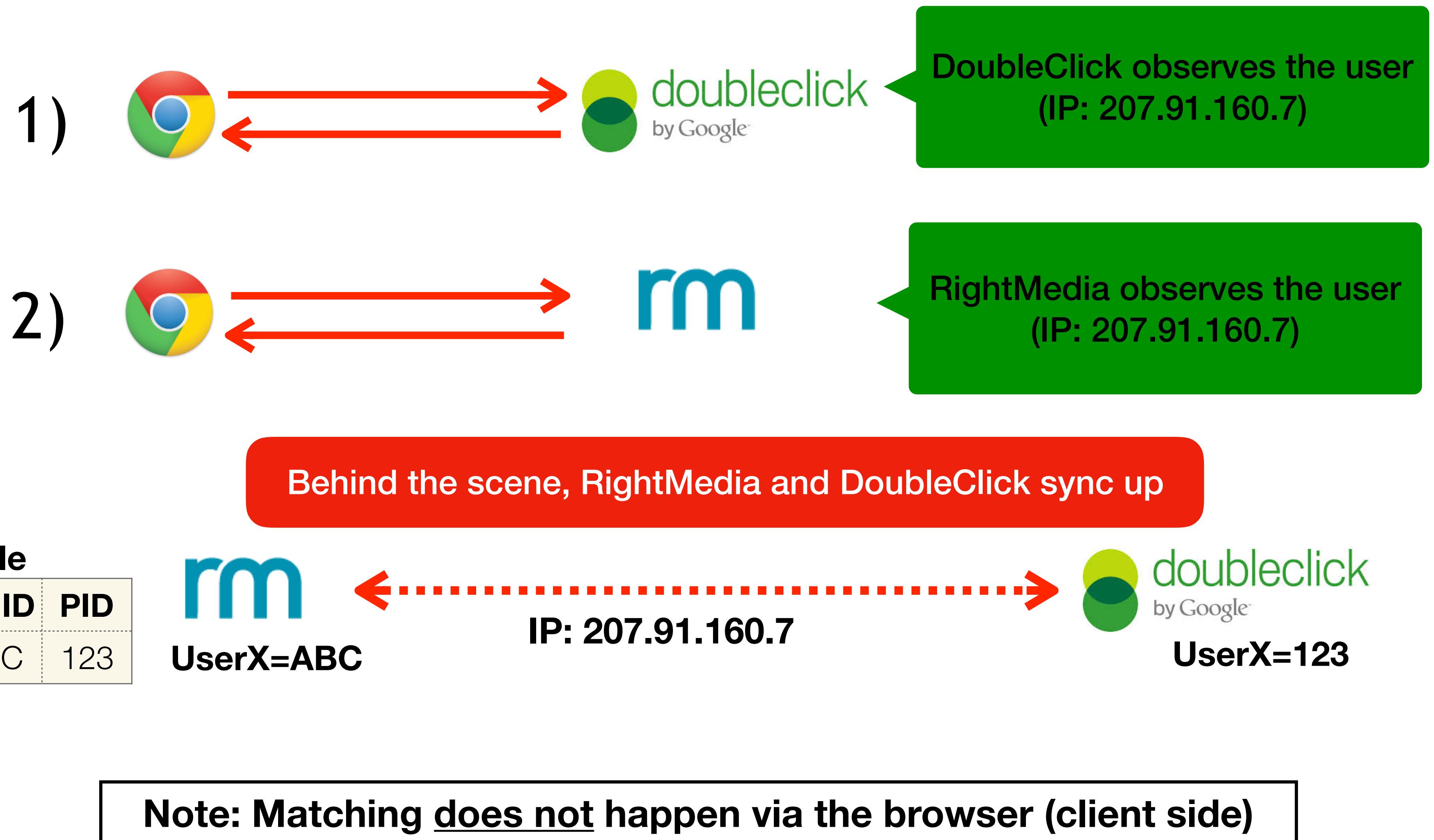
Limitation #1: Server Side Matching



Limitation #1: Server Side Matching



Limitation #1: Server Side Matching



Limitation #2: Obfuscation

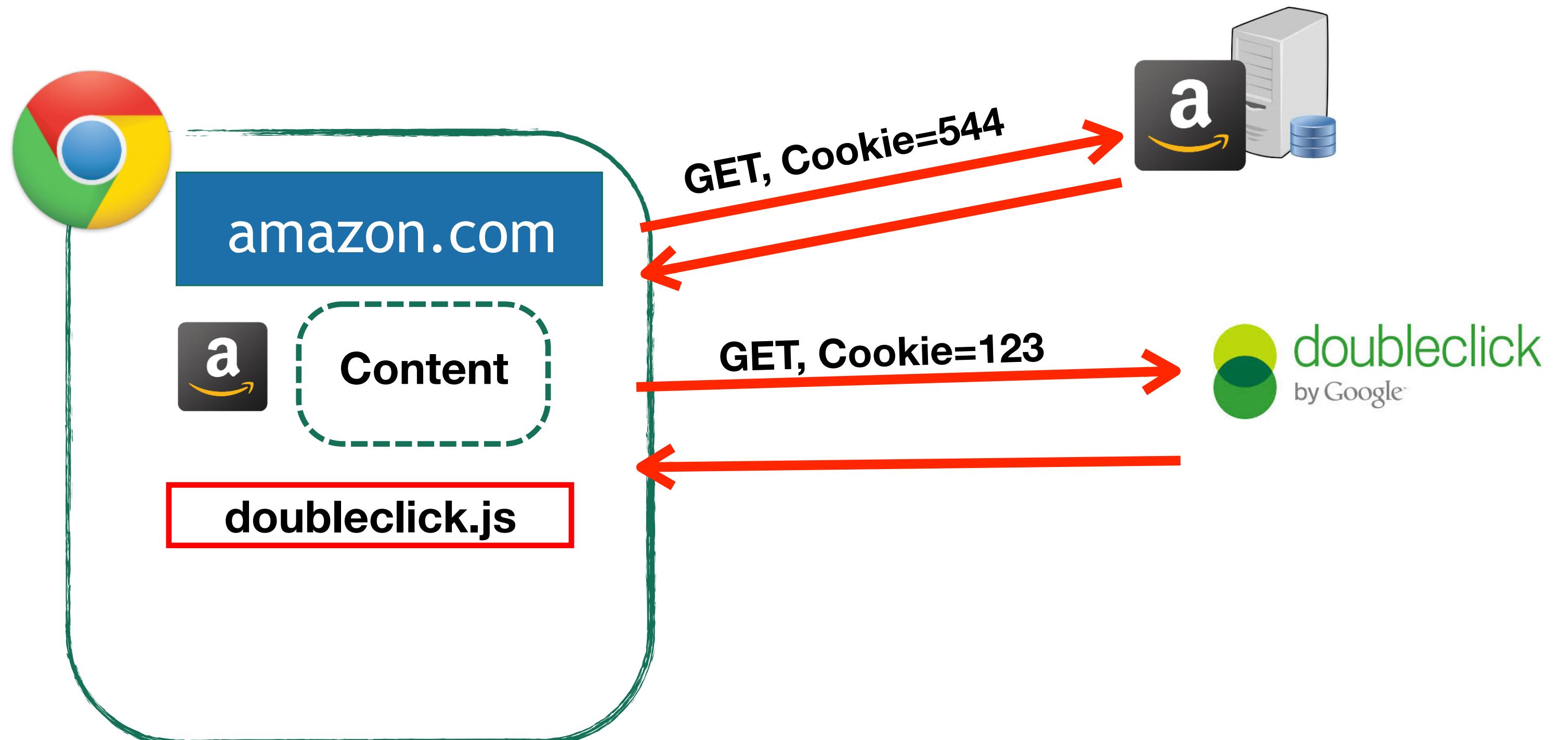
Limitation #2: Obfuscation



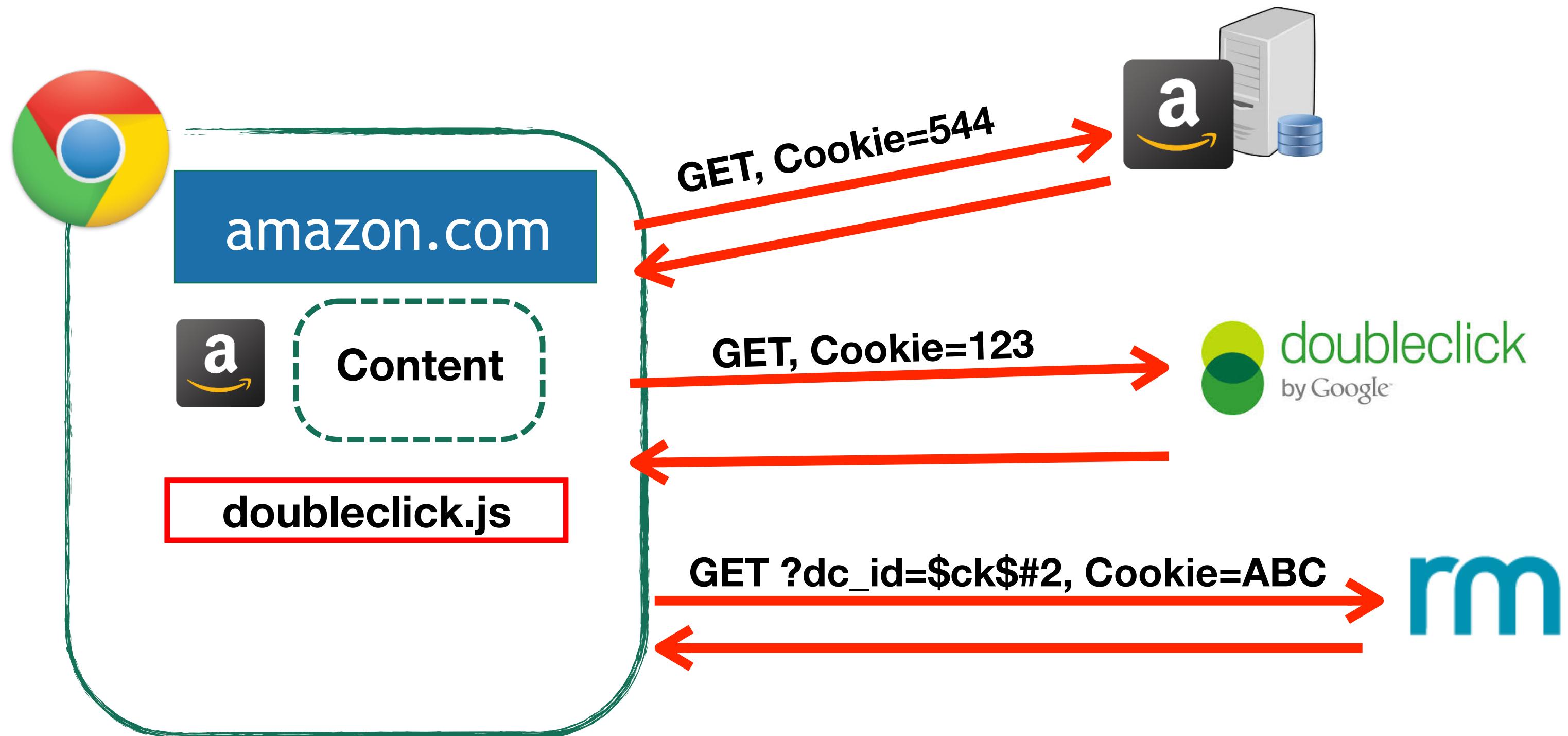
Limitation #2: Obfuscation



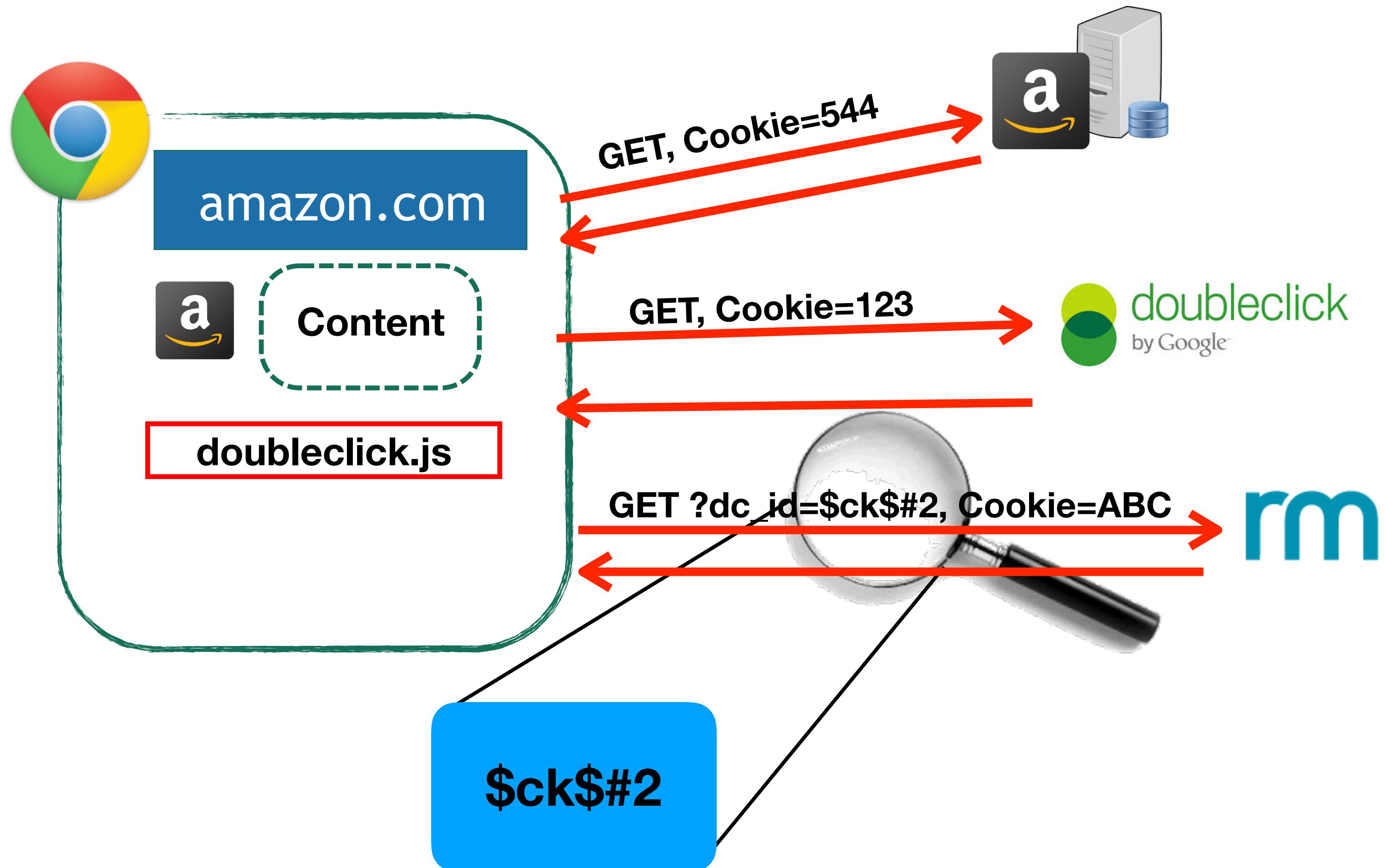
Limitation #2: Obfuscation



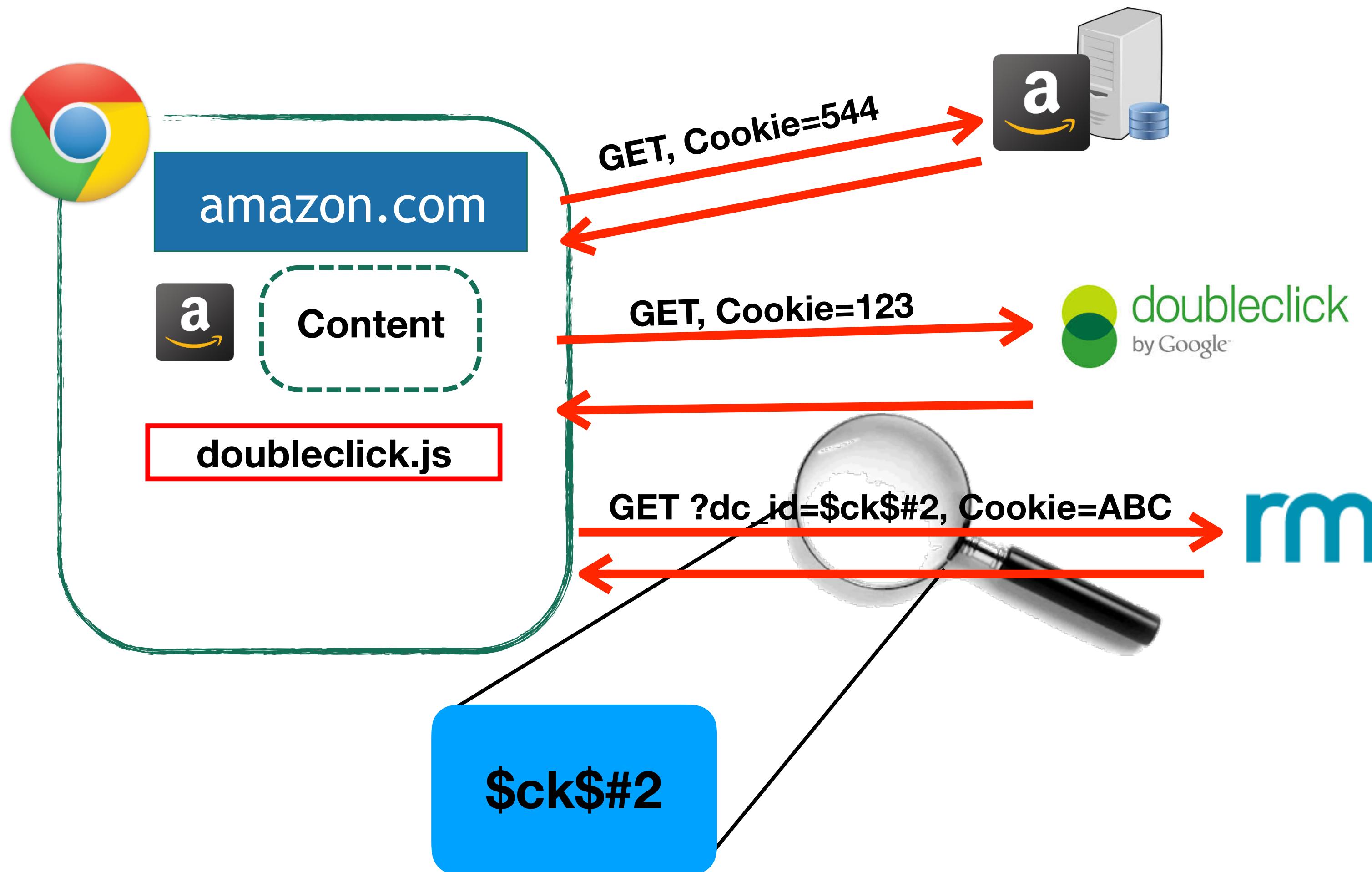
Limitation #2: Obfuscation



Limitation #2: Obfuscation



Limitation #2: Obfuscation

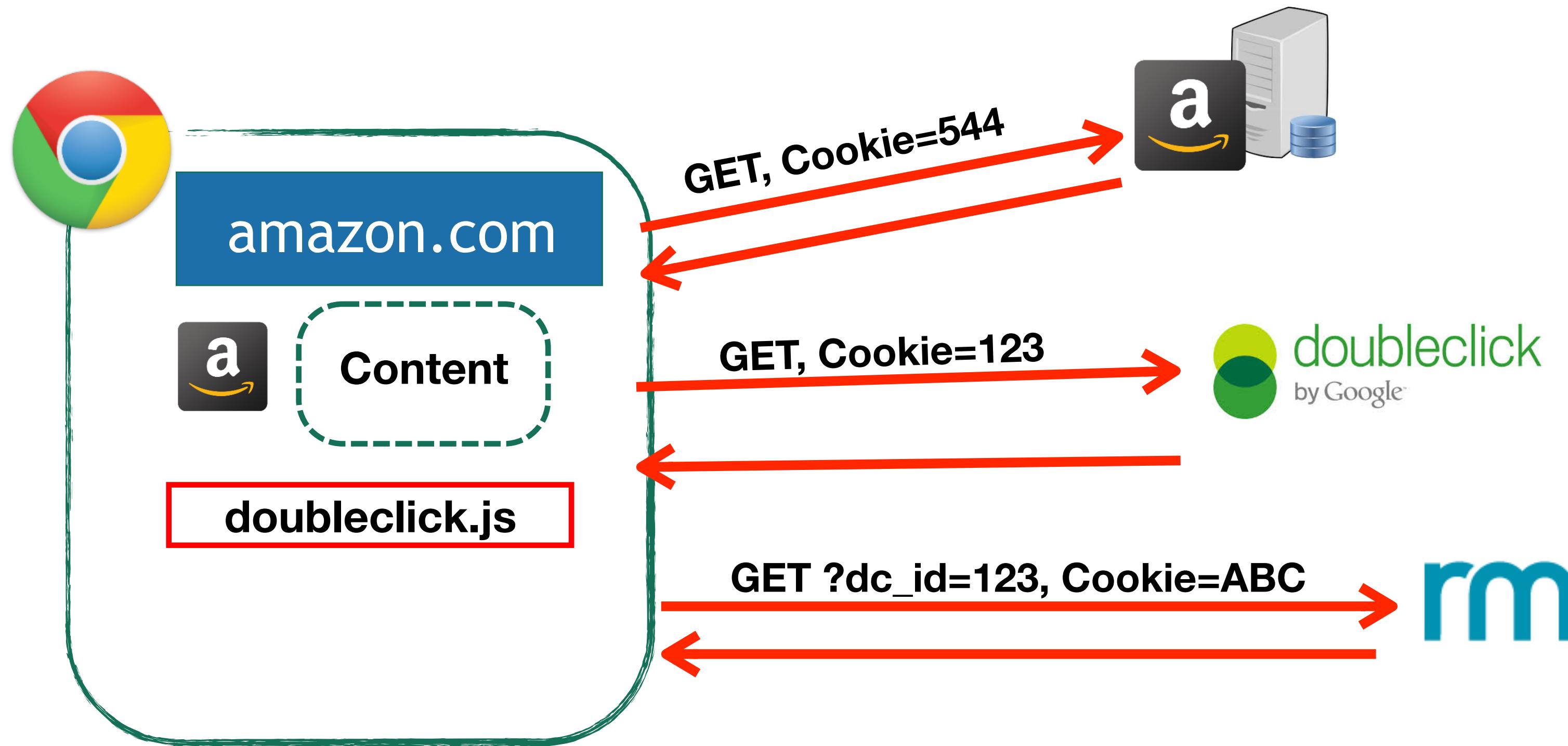


Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

We won't be able to detect Cookie Matching using HTTP headers

Limitation #3: Attribution of Resources



Matching Table

Partner	My ID	PID
DoubleClick	ABC	123

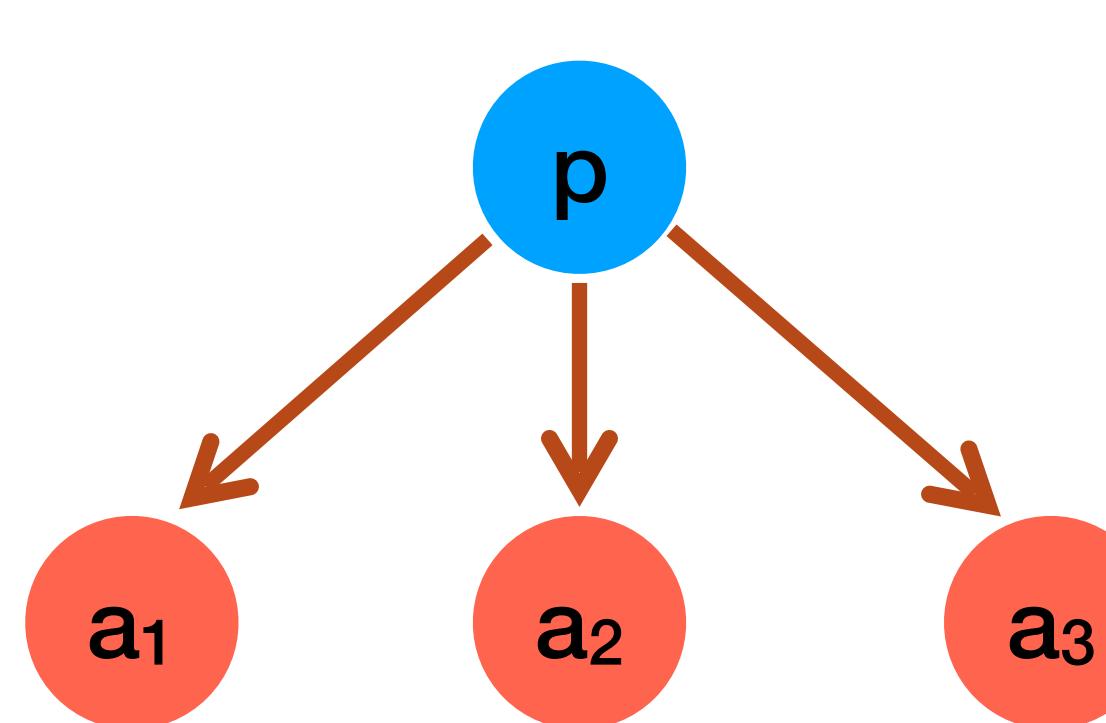
Resource Attribution

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



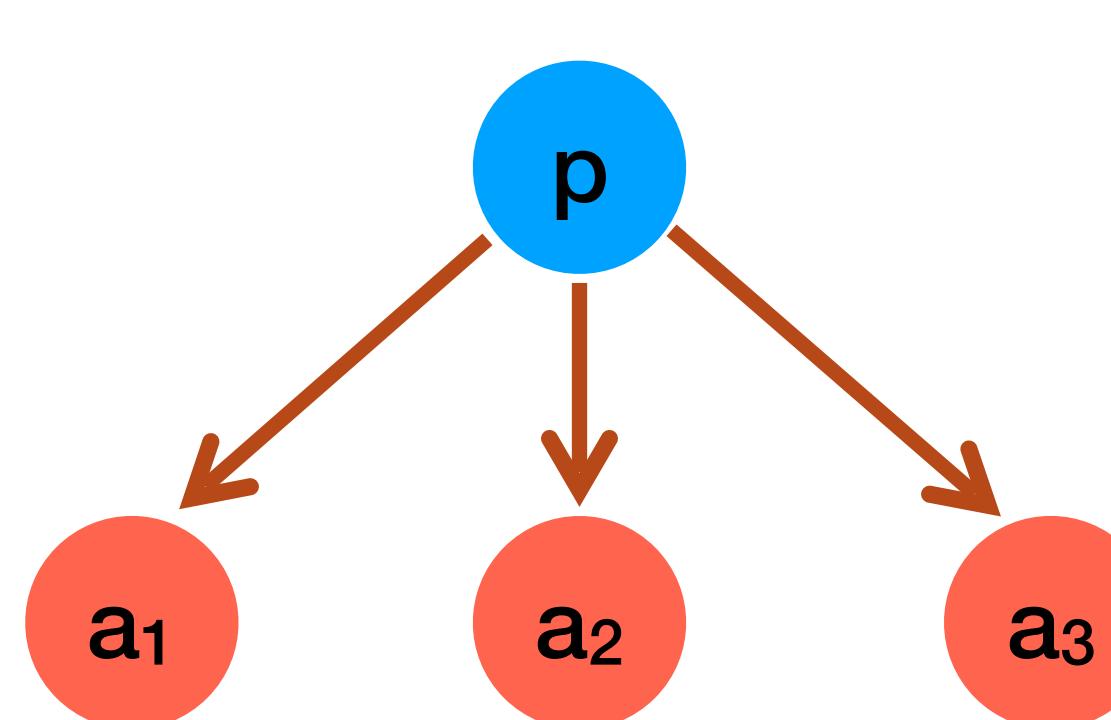
Resource Attribution

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
         by cookie-match.js --&gt;
    &lt;img src="a2.com/pixel.jpg"/&gt;

    &lt;iframe src="a3.com/banner.html"&gt;
      &lt;script src="a4.com/ads.js"&gt;&lt;/script&gt;
    &lt;/iframe&gt;
  &lt;/body&gt;
&lt;/html&gt;</pre>
```

Attribution with DOM



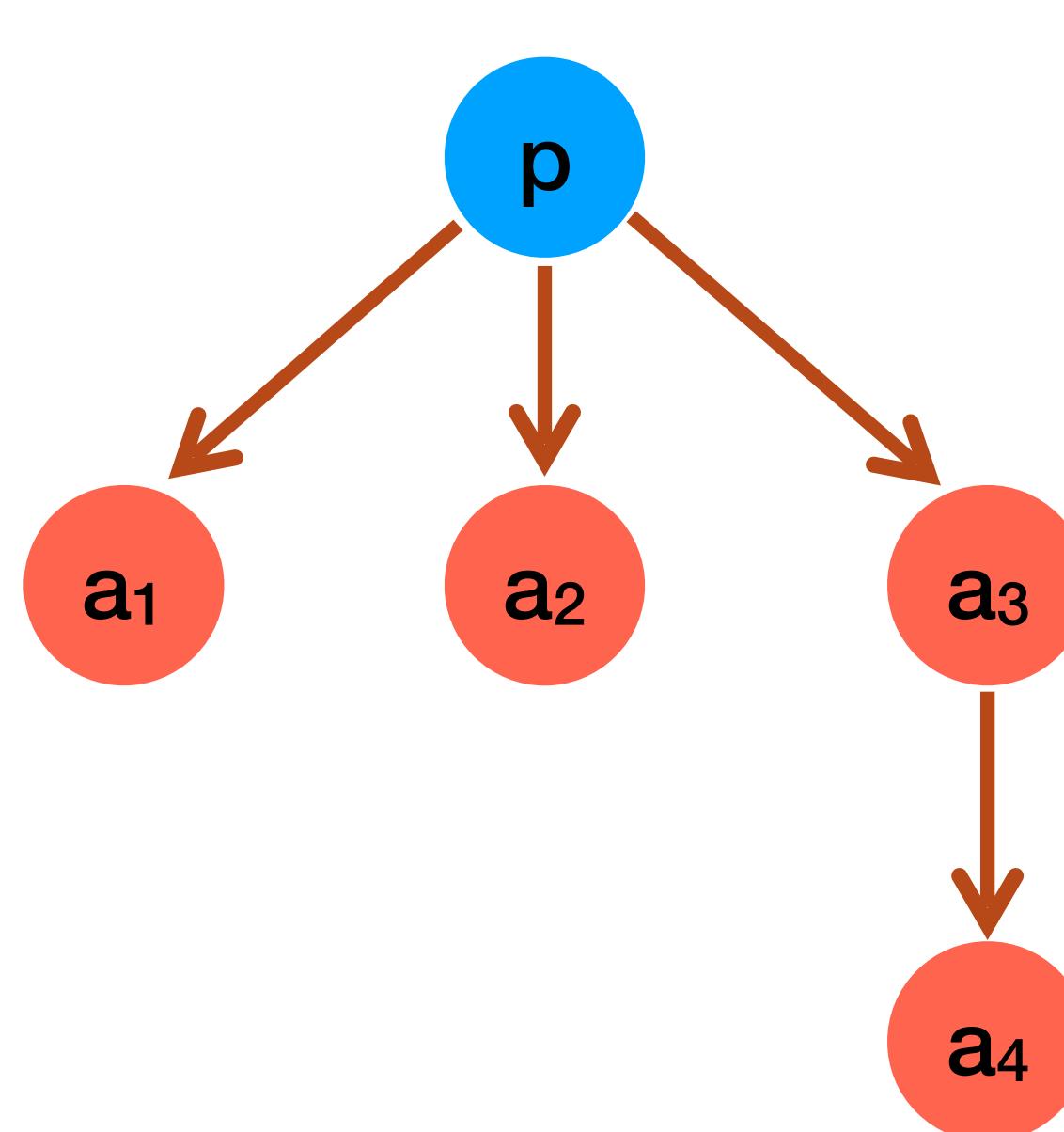
Resource Attribution

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



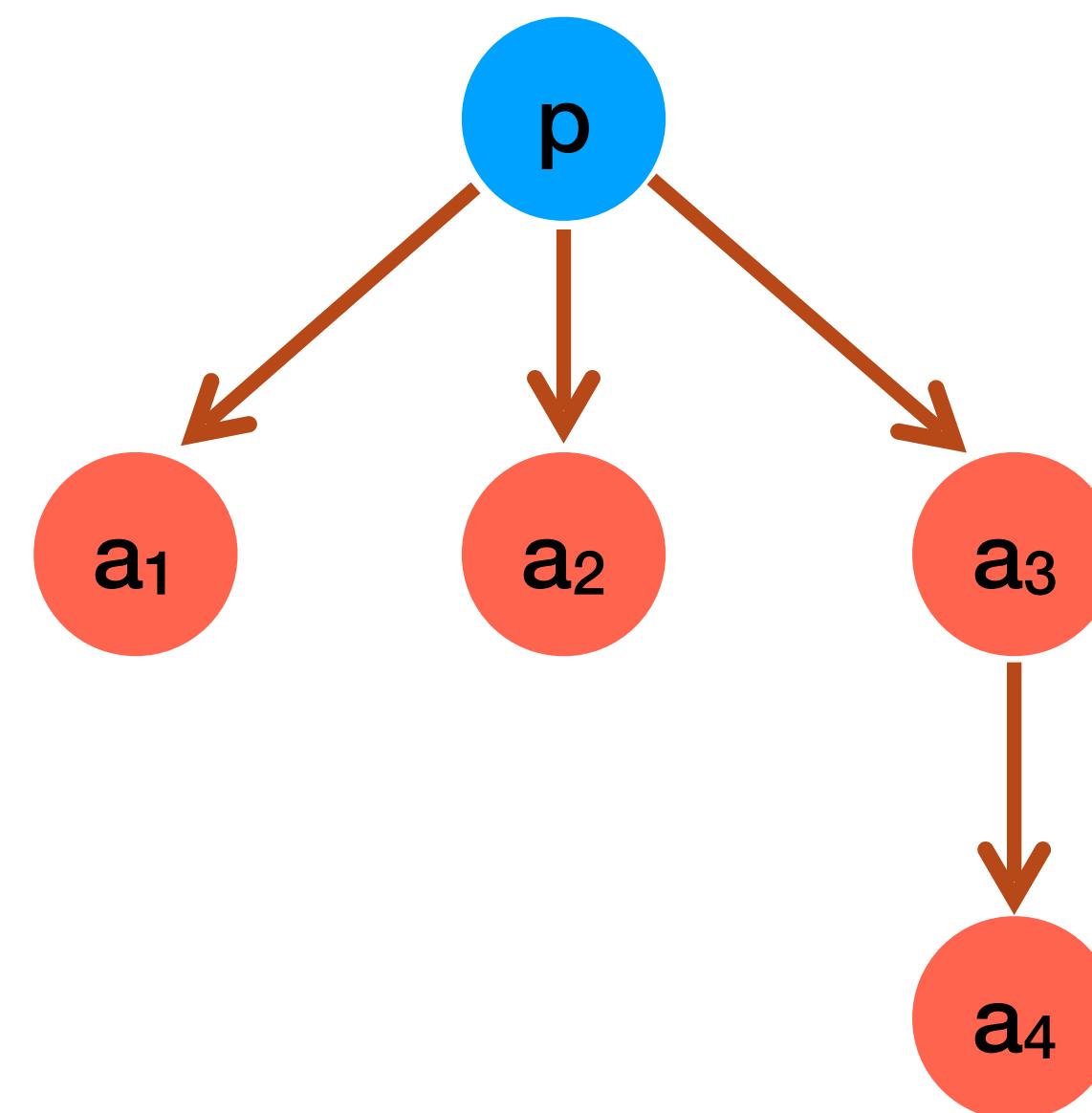
Resource Attribution

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



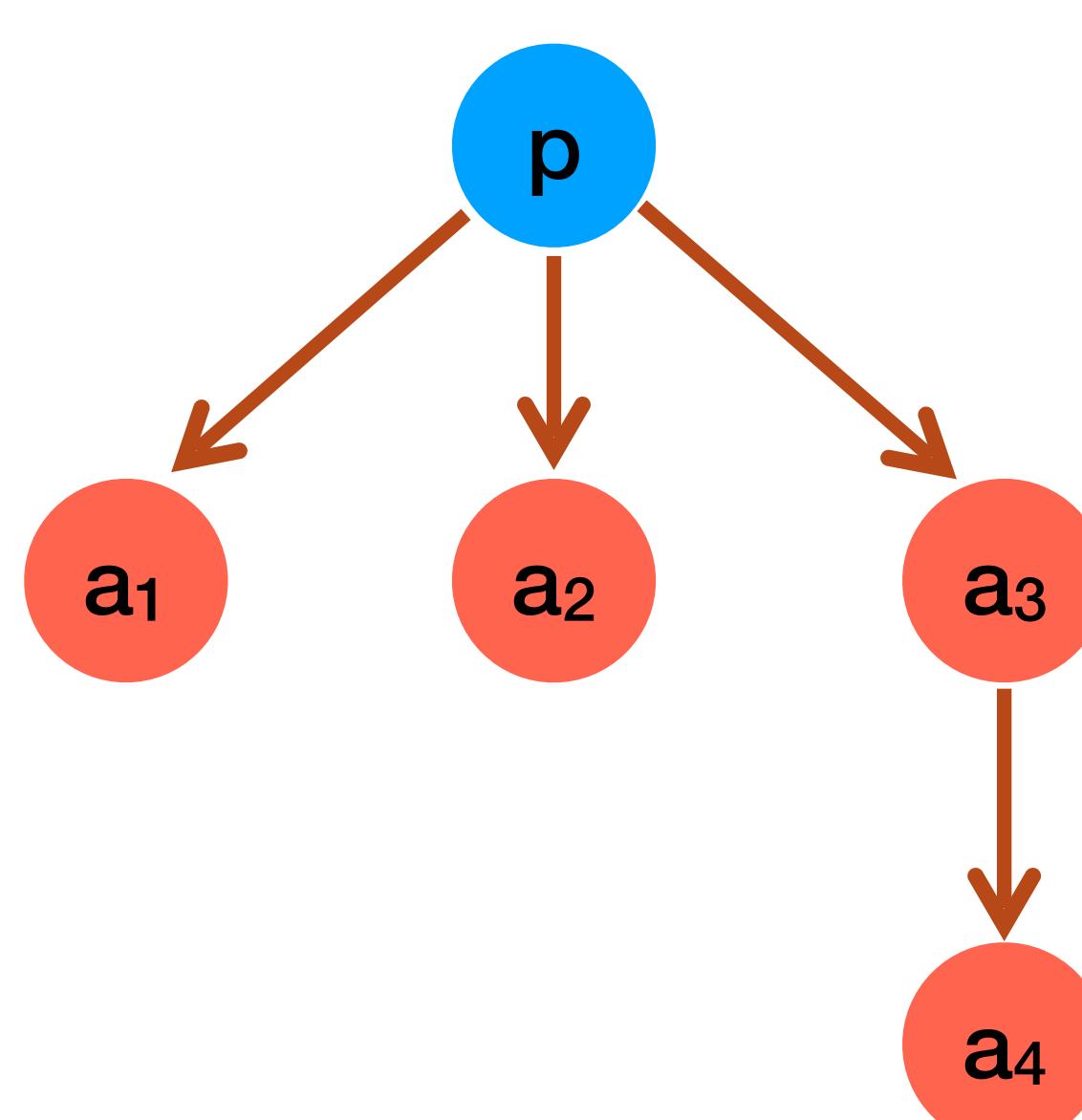
Resource Attribution

DOM Tree for <http://p.com/index.html>

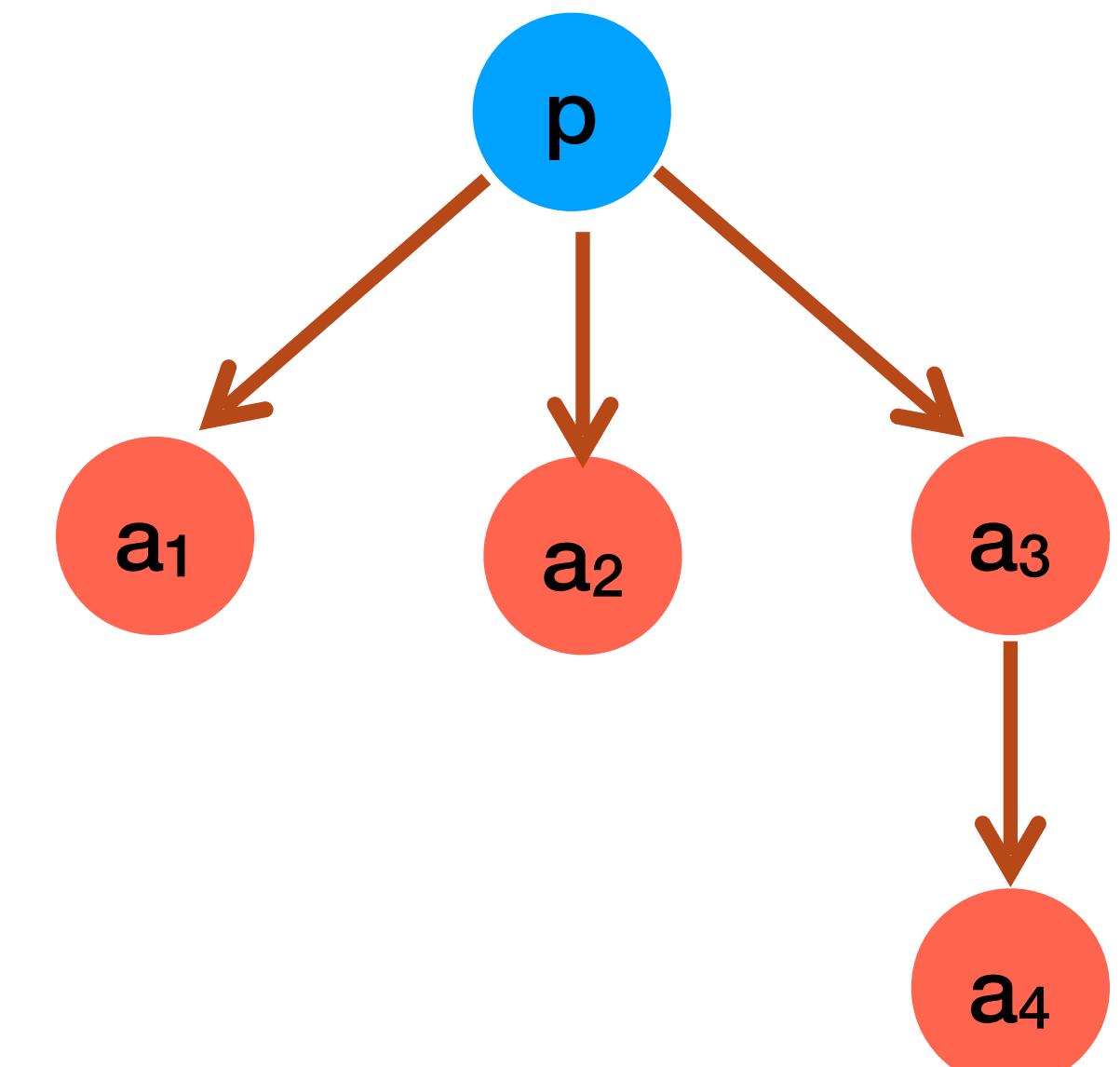
```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



Correct Attribution



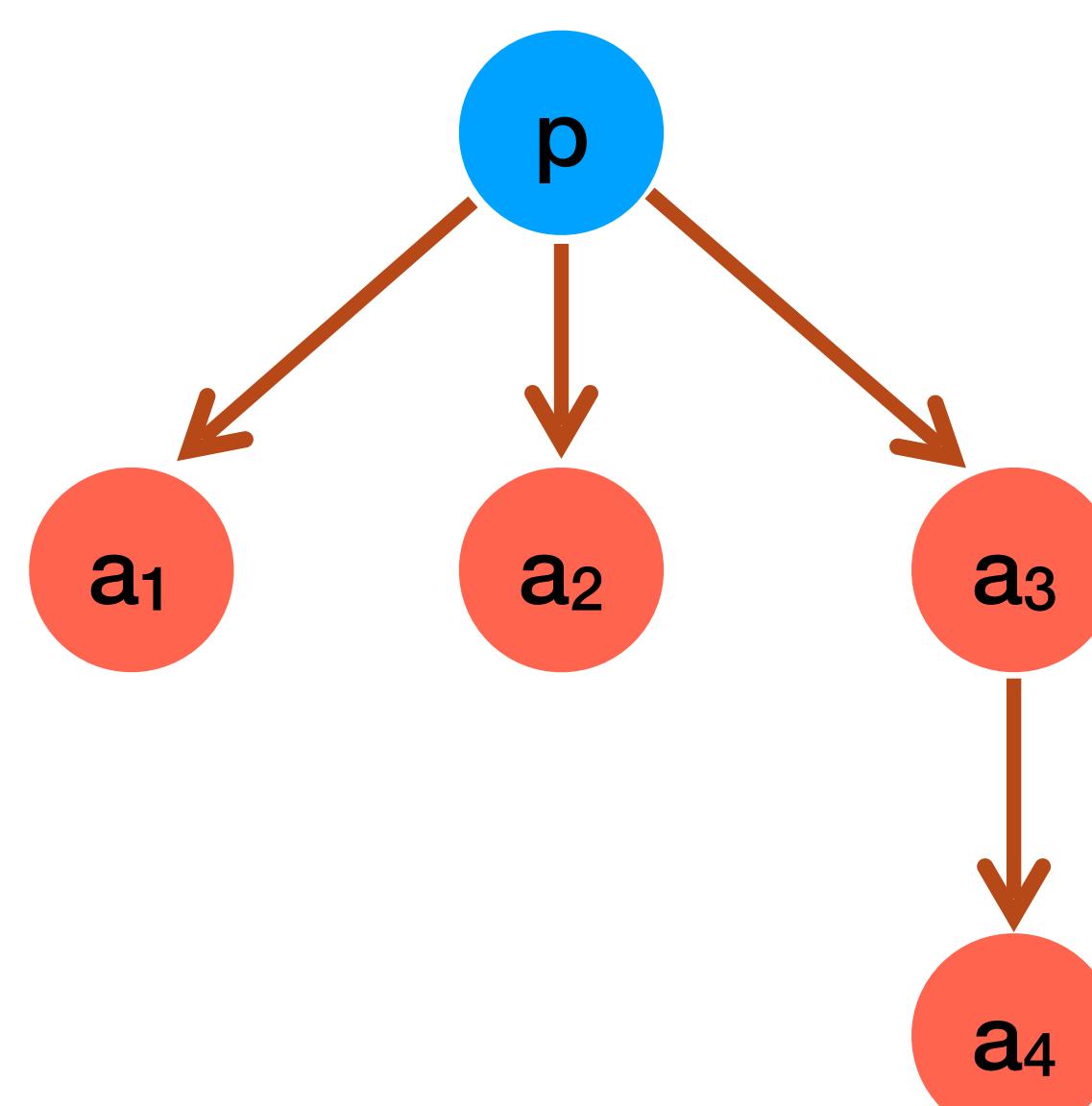
Resource Attribution

DOM Tree for <http://p.com/index.html>

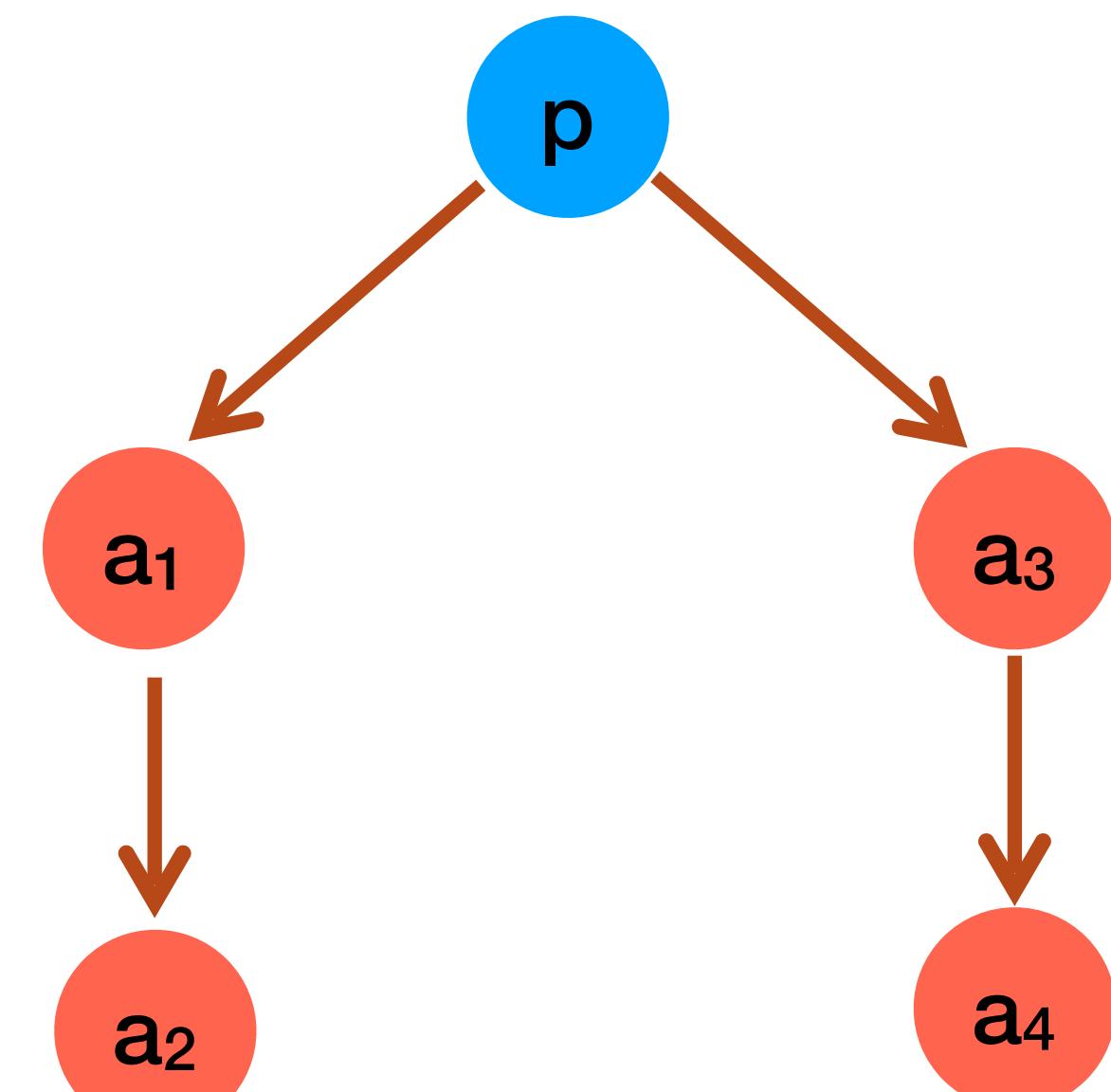
```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



Correct Attribution



Resource Attribution

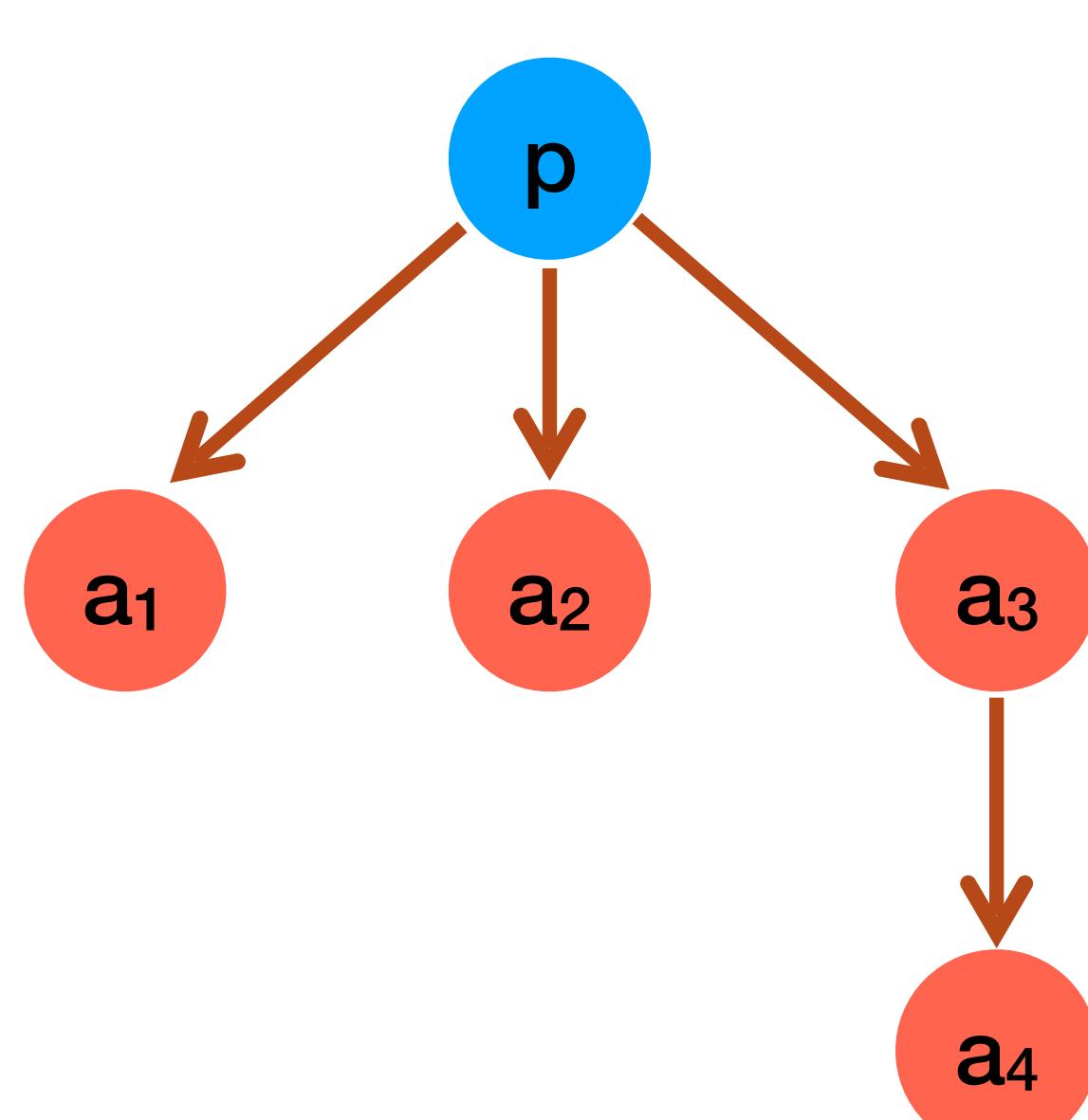
Instrumented Chromium¹ binary that records the provenance of page elements

DOM Tree for <http://p.com/index.html>

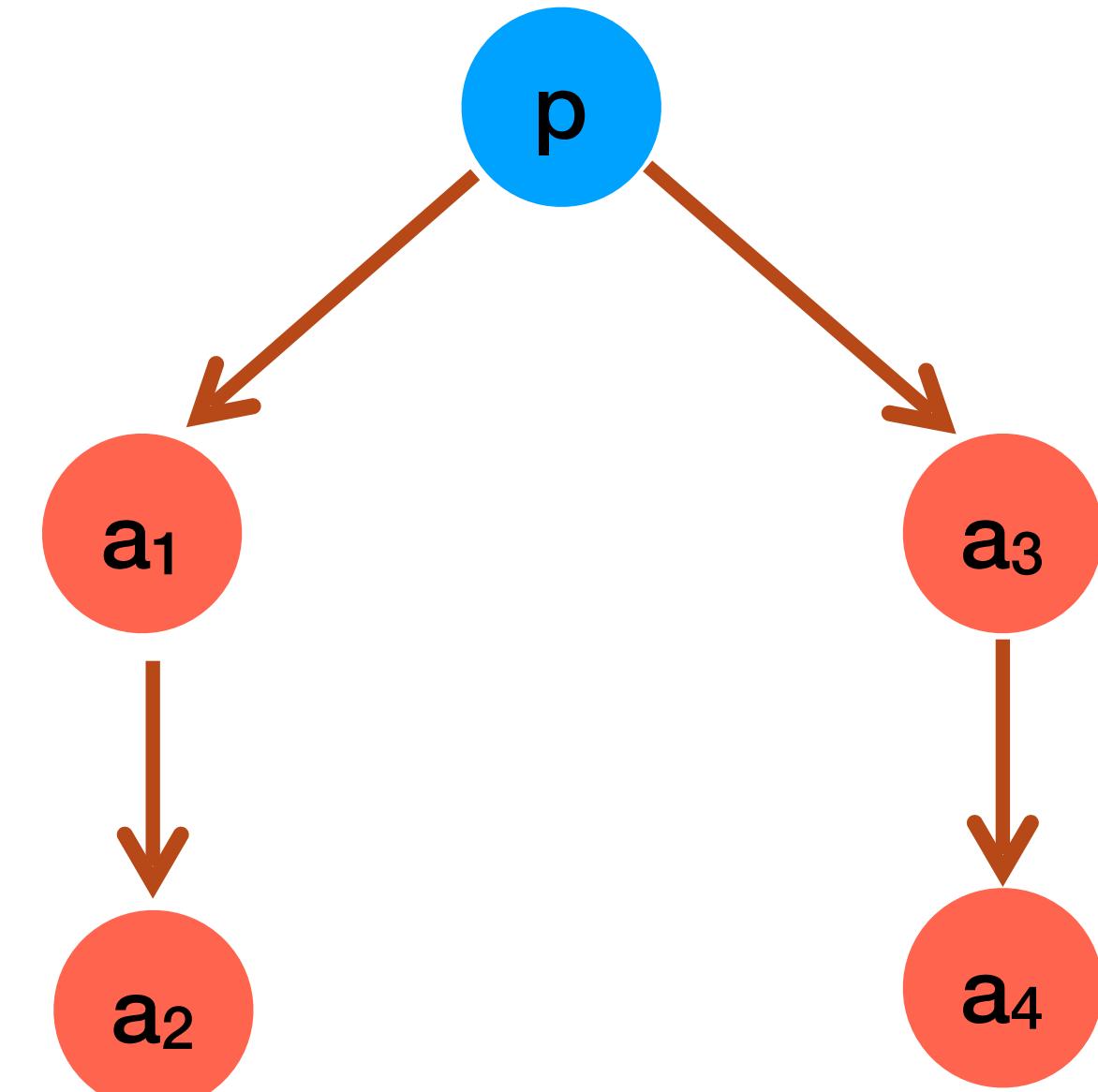
```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Attribution with DOM



Correct Attribution

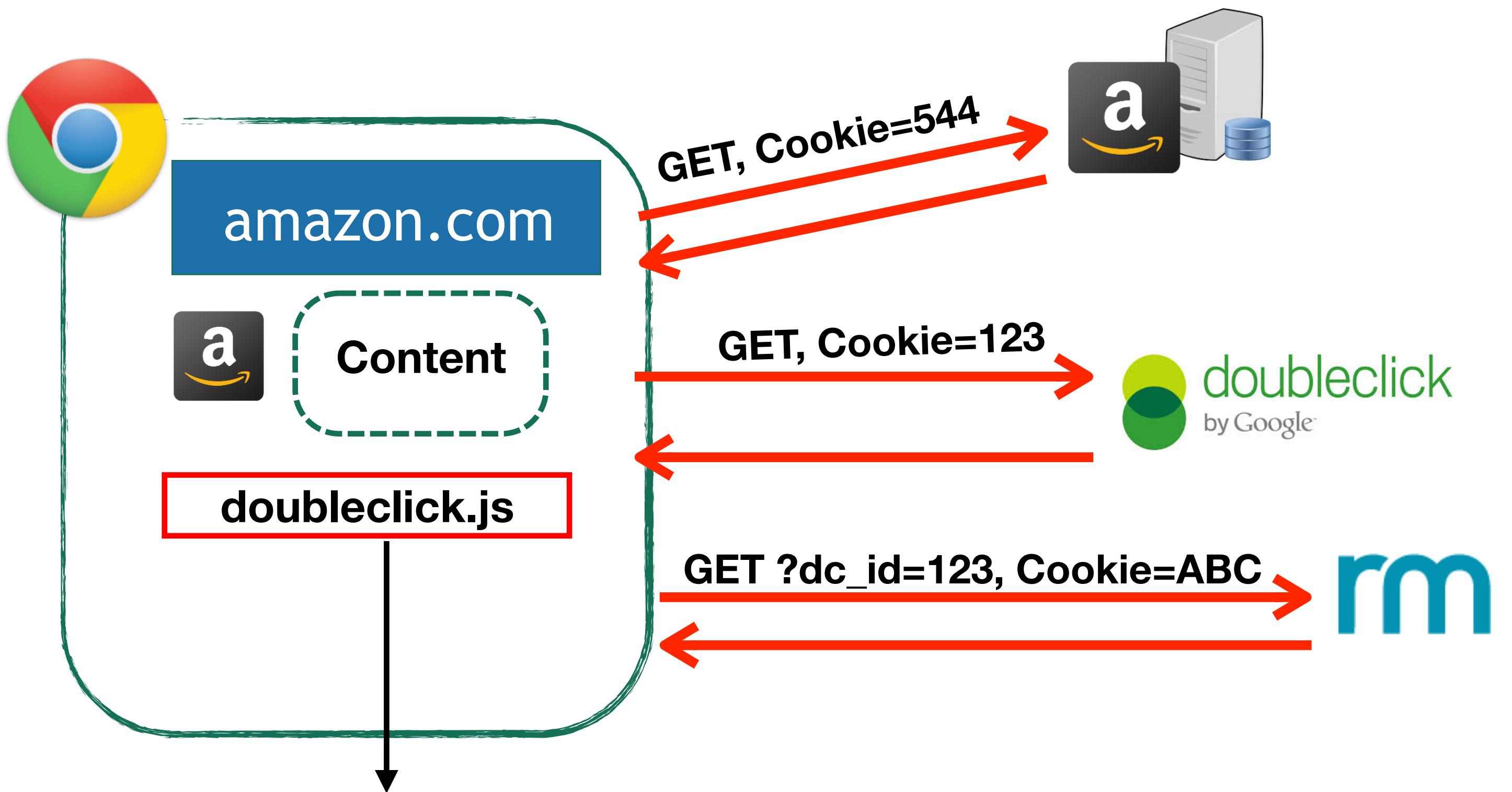


1: Arshad et al : Include Me Out: In-Browser Detection of Malicious Third-Party Content Inclusions, FC 2016

Limitation #3: Attribution of Resources

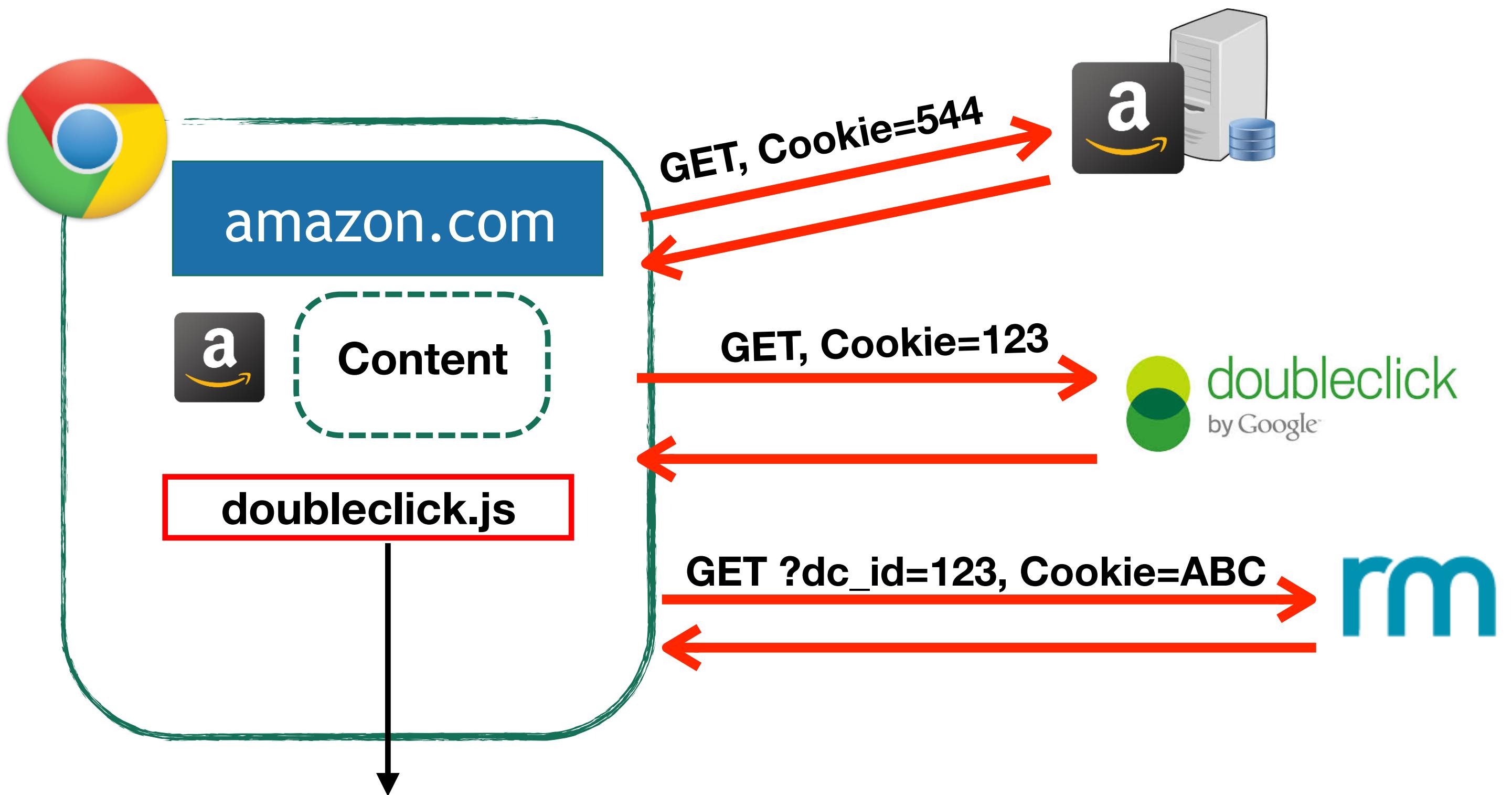


Limitation #3: Attribution of Resources



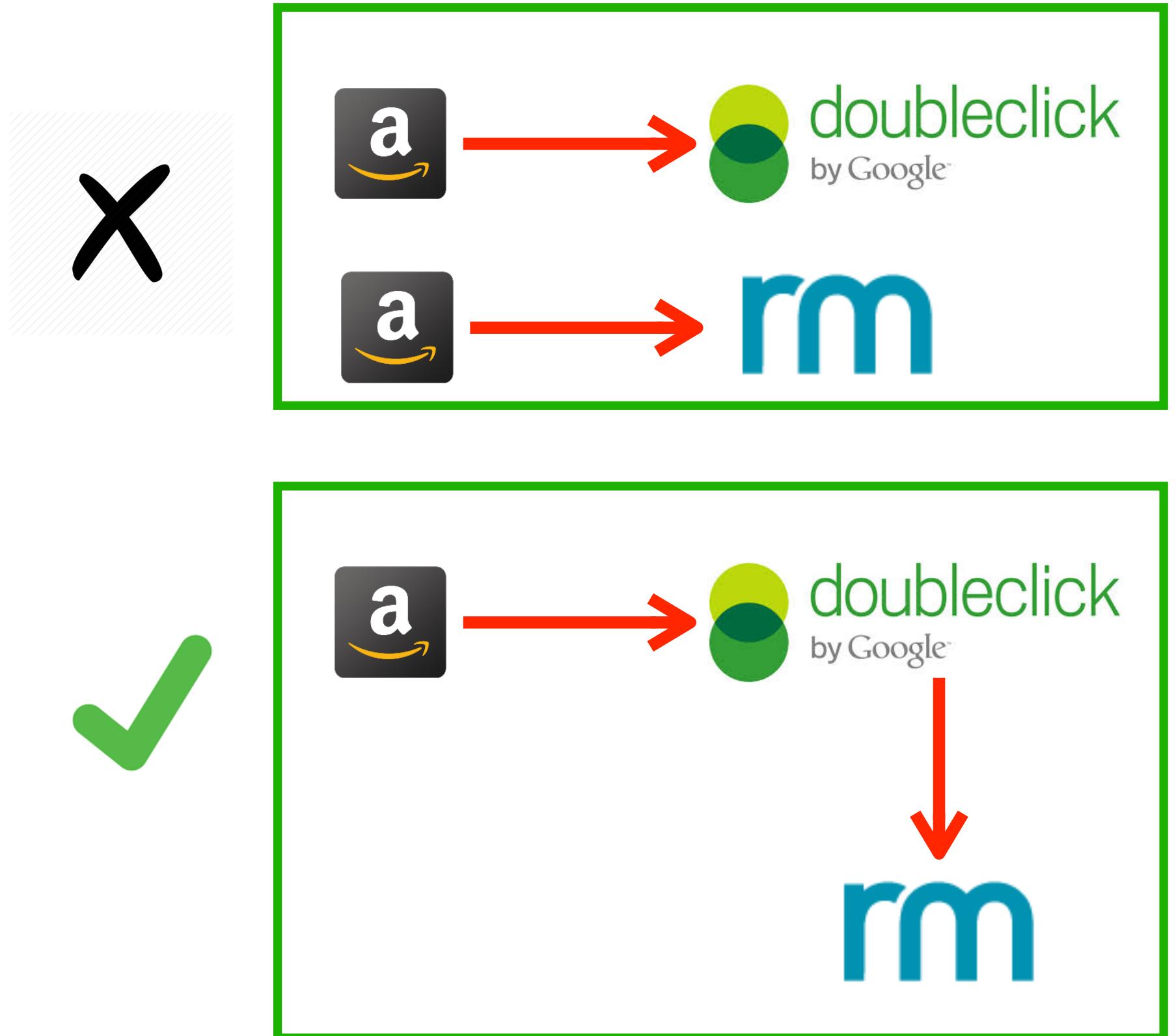
Since JavaScript from DoubleClick is in the context of the main page, the **Referer will be Amazon**

Limitation #3: Attribution of Resources



Since JavaScript from DoubleClick is in the context of the main page, the **Referer will be Amazon**

Limitation #3: Attribution of Resources



Since JavaScript from DoubleClick is in the context of the main page, the **Referer will be Amazon**

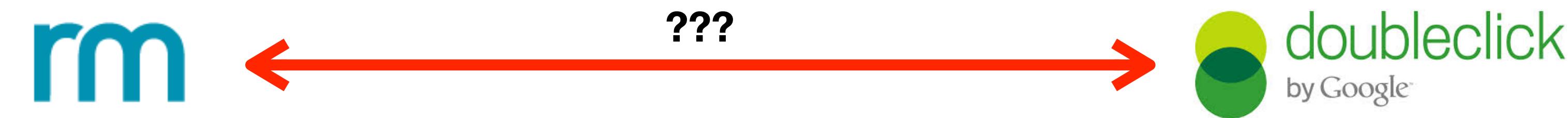
Limitation #3: Attribution of Resources



Since JavaScript from DoubleClick is in the context of the main page, the **Referer will be Amazon**

Referer value in HTTP header can lead to incorrect attribution

A New Approach To Detect Information Sharing



1. Resilient to obfuscation

- Should be able to detect matching without relying on content.

2. Platform agnostic

- Should be able to detect both client-side (browser) and server-side matching.

3. Provides strong attribution

- Should be able to detect who initiated matching and with whom.

Using Retargeted Ads as a Detection Mechanism

Retargeted ads: The ads which follow you around on the web

- Highly targeted (specific) ads
- Very expensive: around \$1 per impression vs. >\$0.01 for contextual ads

Using Retargeted Ads as a Detection Mechanism

Retargeted ads: The ads which follow you around on the web

- Highly targeted (specific) ads
- Very expensive: around \$1 per impression vs. >\$0.01 for contextual ads



Cisco-Linksys AE1000 High-Performance Wireless-N Adapter
by Linksys
 207 customer reviews | 10 answered questions
Price: **\$15.99**
Only 1 left in stock.
Want it Tuesday, June 14? Order within 33 hrs 50 mins and choose One-Day Shipping at checkout. [Details](#)
Sold by Home Sweet Home Direct and Fulfilled by Amazon.
Eligible for [amazon smile](#) donation.

 Want to hire a computer technician?
Buy professional computer technician services directly on Amazon. Backed by our Happiness Guarantee.
[Learn more](#)
Networking Equipment Features: WEP Security, WPA Security, Easy Setup, WPA2

amazon.com [Shop now](#)
Linksys/Cisco AE1000 300Mbps 802.11n Dual-Band Wireless...
\$39.99 \$15.99
Privacy

Key Insight: Because retargeted ads are so specific, they can be used to conduct controlled experiments

- Information must be shared between ad networks to serve retargeted ads

Retargeted Ads are Very Specific (A Simple Experiment)

1)

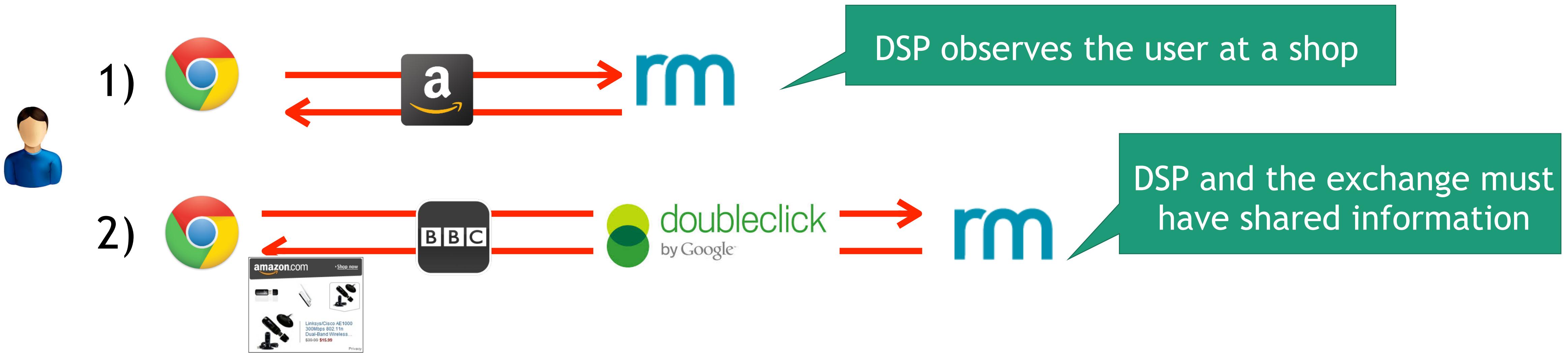


Retargeted Ads are Very Specific (A Simple Experiment)



1. User visits where (DSP) is present and observes the user.

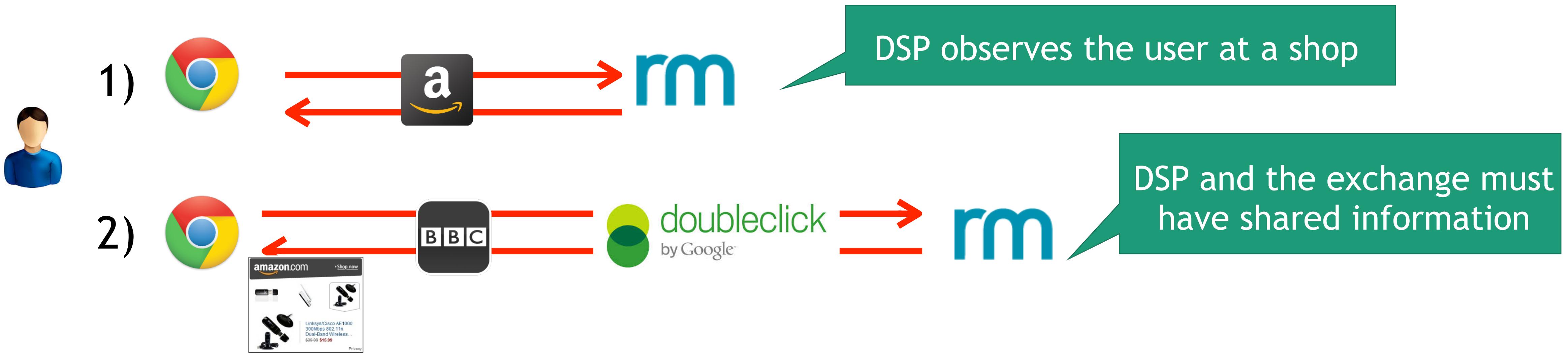
Retargeted Ads are Very Specific (A Simple Experiment)



1. User visits where **rm** (DSP) is present and observes the user.
2. User visits and sees a retargeted ad for Amazon product.



Retargeted Ads are Very Specific (A Simple Experiment)



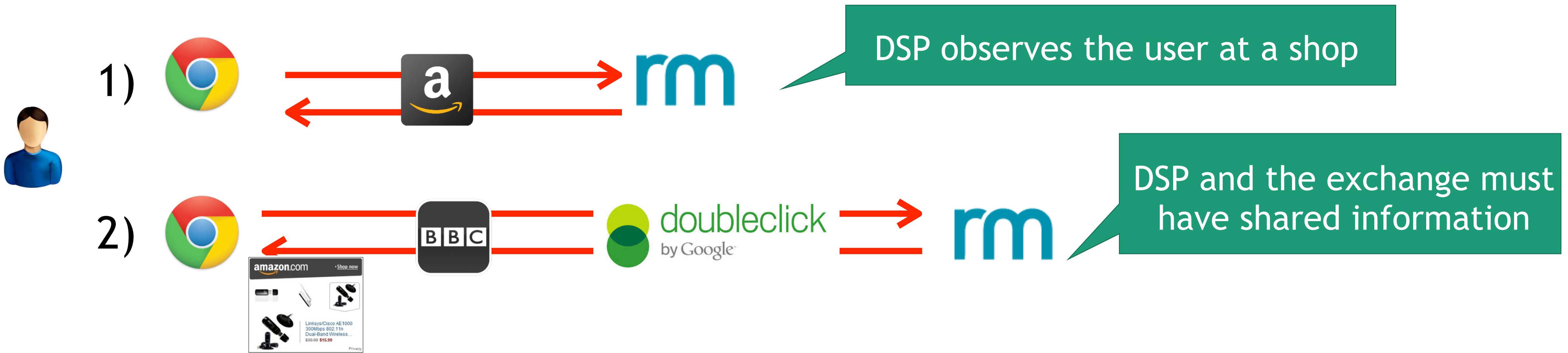
1. User visits **a** where **rm** (DSP) is present and observes the user.
2. User visits **BBC** and sees a retargeted ad for Amazon product.



How does **rm** identify the user on **BBC**, to show a specific ad?

- Retargets are too expensive to be served to random, unknown users

Retargeted Ads are Very Specific (A Simple Experiment)



1. User visits where (DSP) is present and observes the user.
2. User visits and sees a retargeted ad for Amazon product.



How does identify the user on , to show a specific ad?

- Retargets are too expensive to be served to random, unknown users
- This proves a causal flow of information from (exchange) → (DSP)
- **Note:** We do not rely on HTTP headers to determine sharing.

Limitations Addressed

Limitations Addressed

Strong Attribution

Content Agnostic

Platform Agnostic

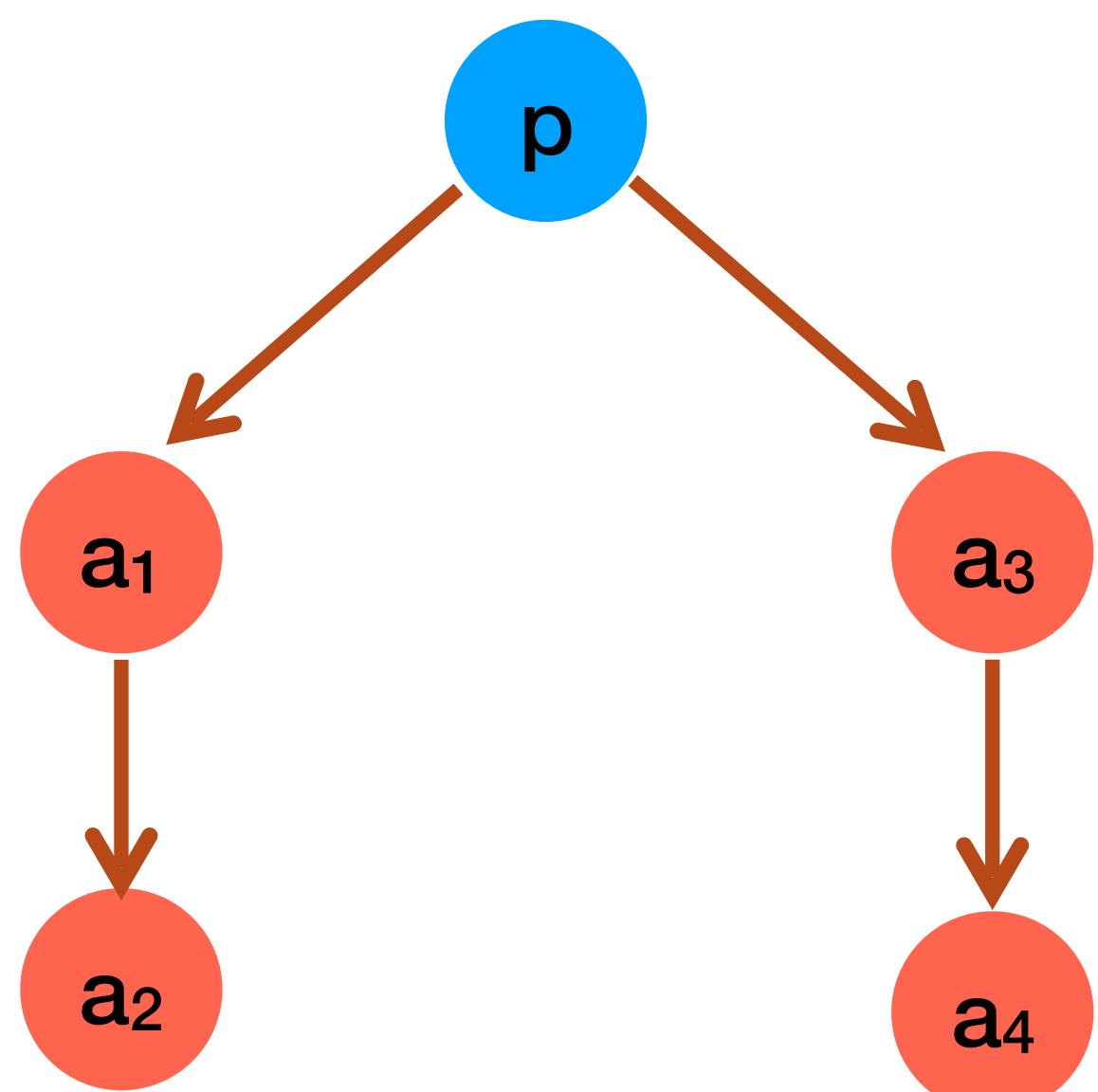
Limitations Addressed

Strong Attribution

Content Agnostic

Platform Agnostic

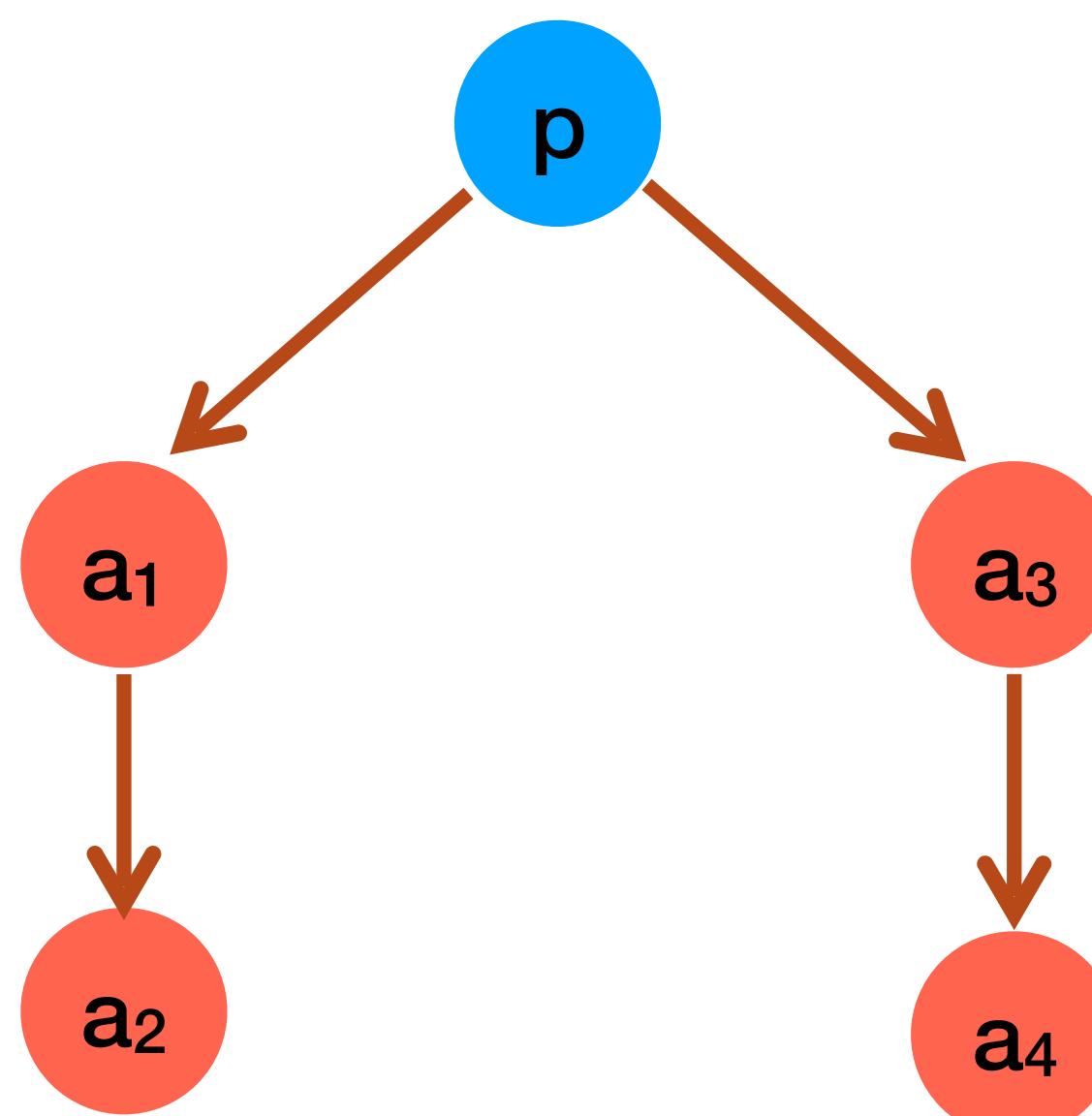
Inclusion Chains



Limitations Addressed

Strong Attribution

Inclusion Chains



Content Agnostic



Cisco-Linksys AE1000 High-Performance Wireless-N Adapter

by [Linksys](#)

 207 customer reviews | 10 answered questions

Price: **\$15.99** 

Only 1 left in stock.

Want it Tuesday, June 14? Order within **33 hrs 50 mins** and choose **One-Day Shipping** at checkout. [Details](#)

Sold by [Home Sweet Home Direct](#) and [Fulfilled by Amazon](#).

Eligible for [amazon smile](#) donation.

 Want to hire a computer technician?

Buy professional computer technician services directly on Amazon. Backed by our Happiness Guarantee.

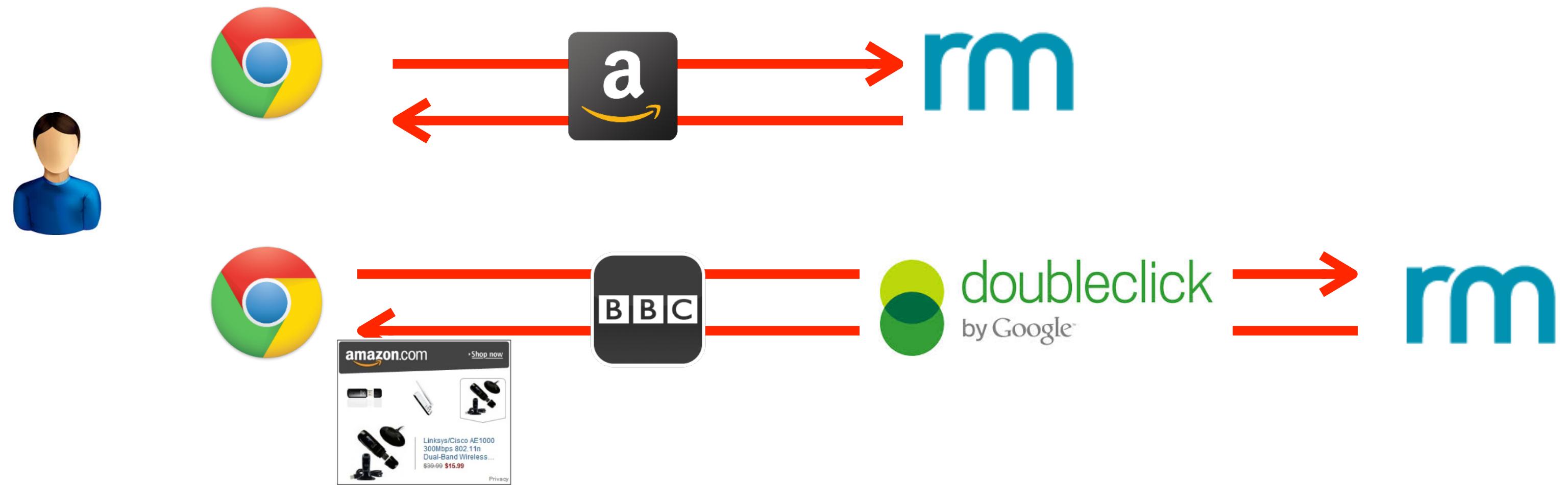
[Learn more](#)

- Networking Equipment Features: WEP Security, WPA Security, Easy Setup, WPA2



Platform Agnostic

The Caveat



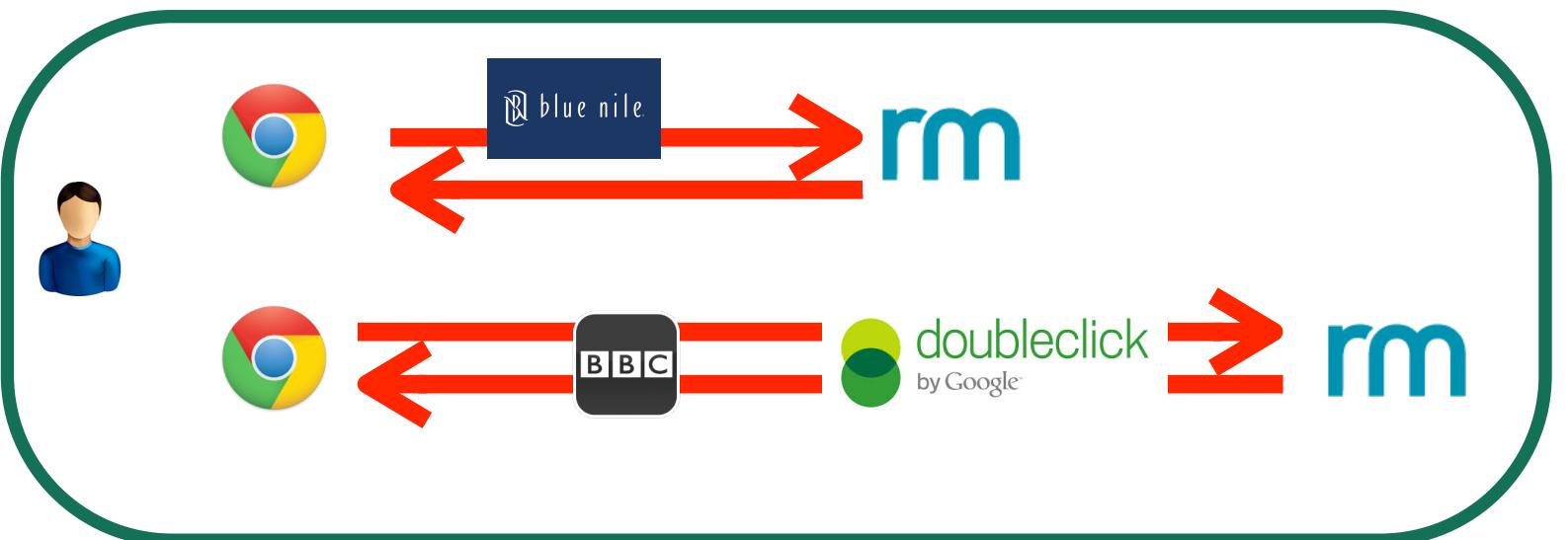
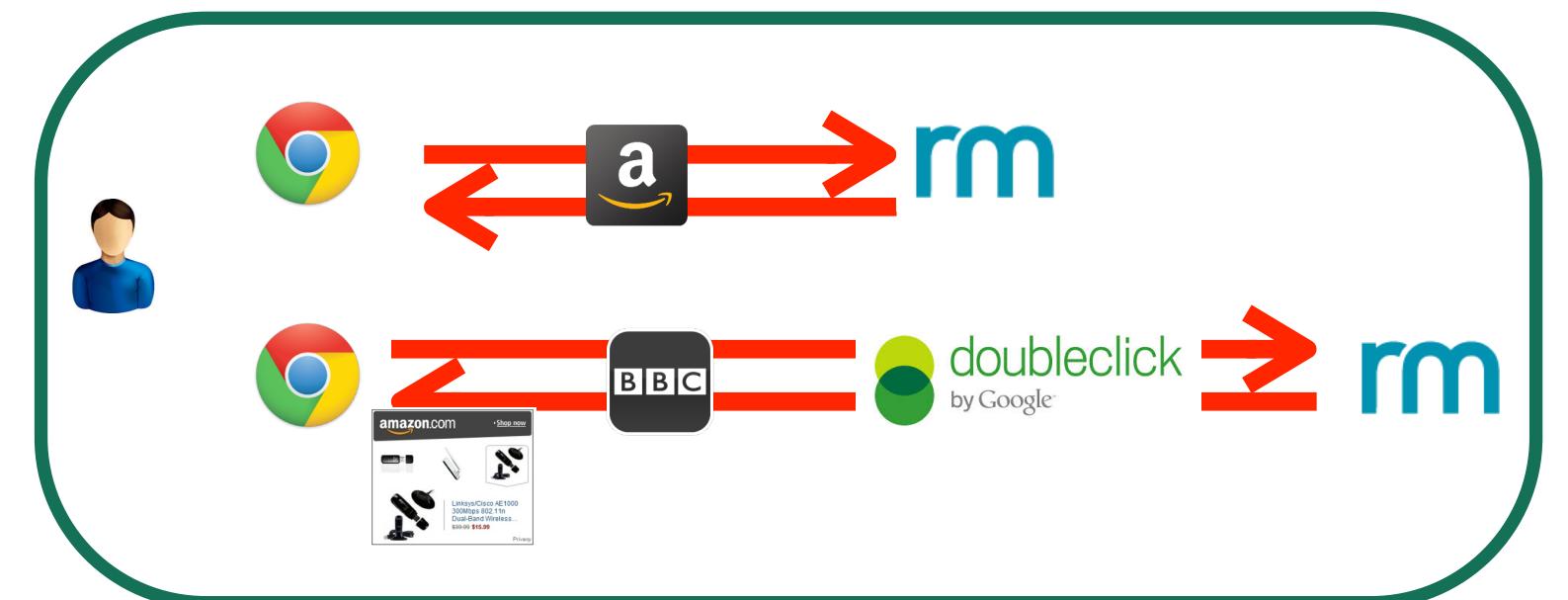
How can we be certain that there is a flow of tracking information from



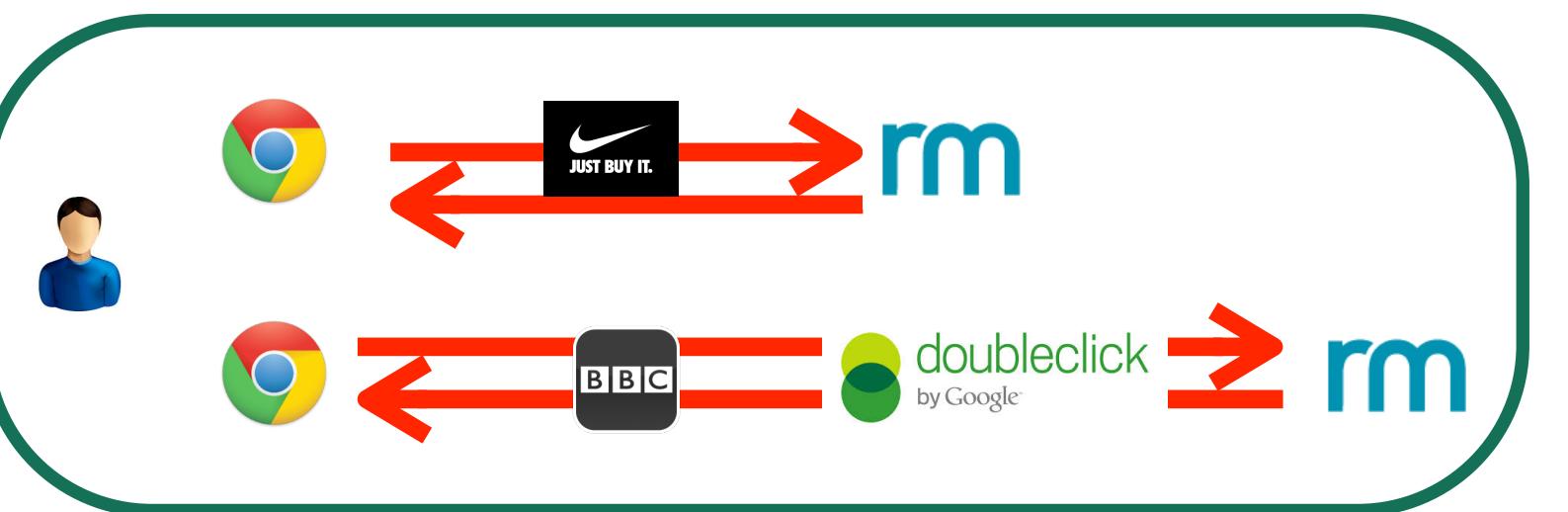
for the ad to be served?

It is possible that amazon is running a very aggressive campaign with a huge budget

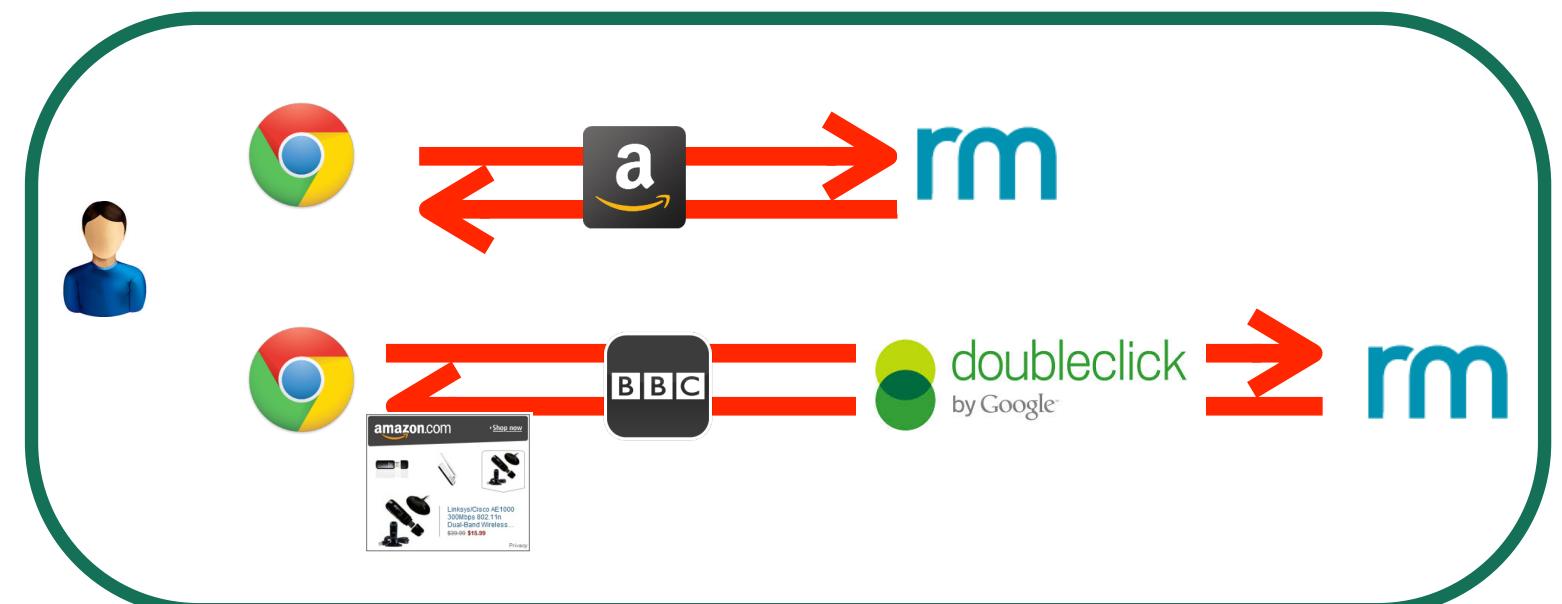
Controlled Experiments



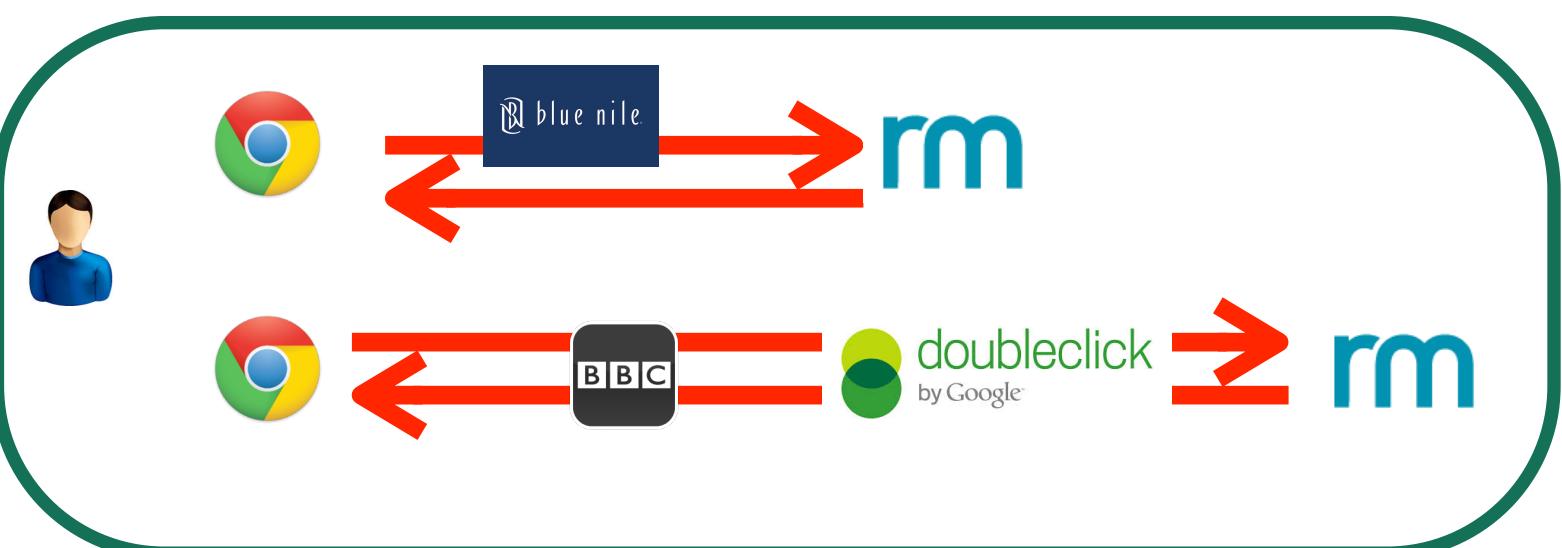
-



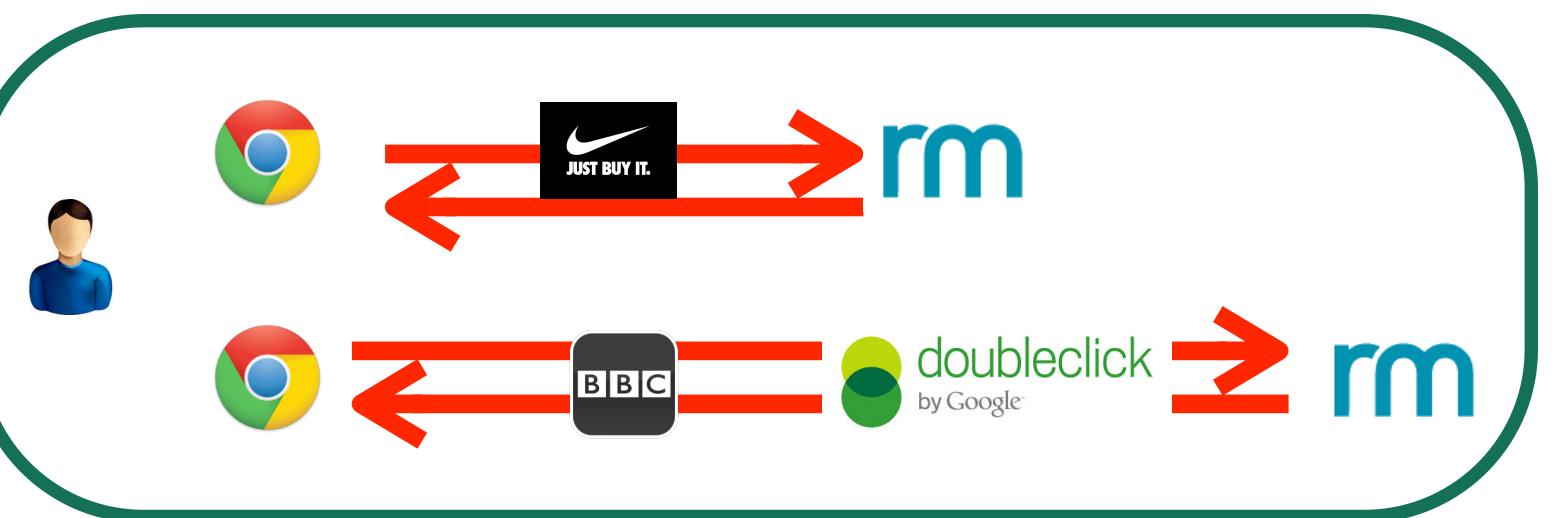
Controlled Experiments



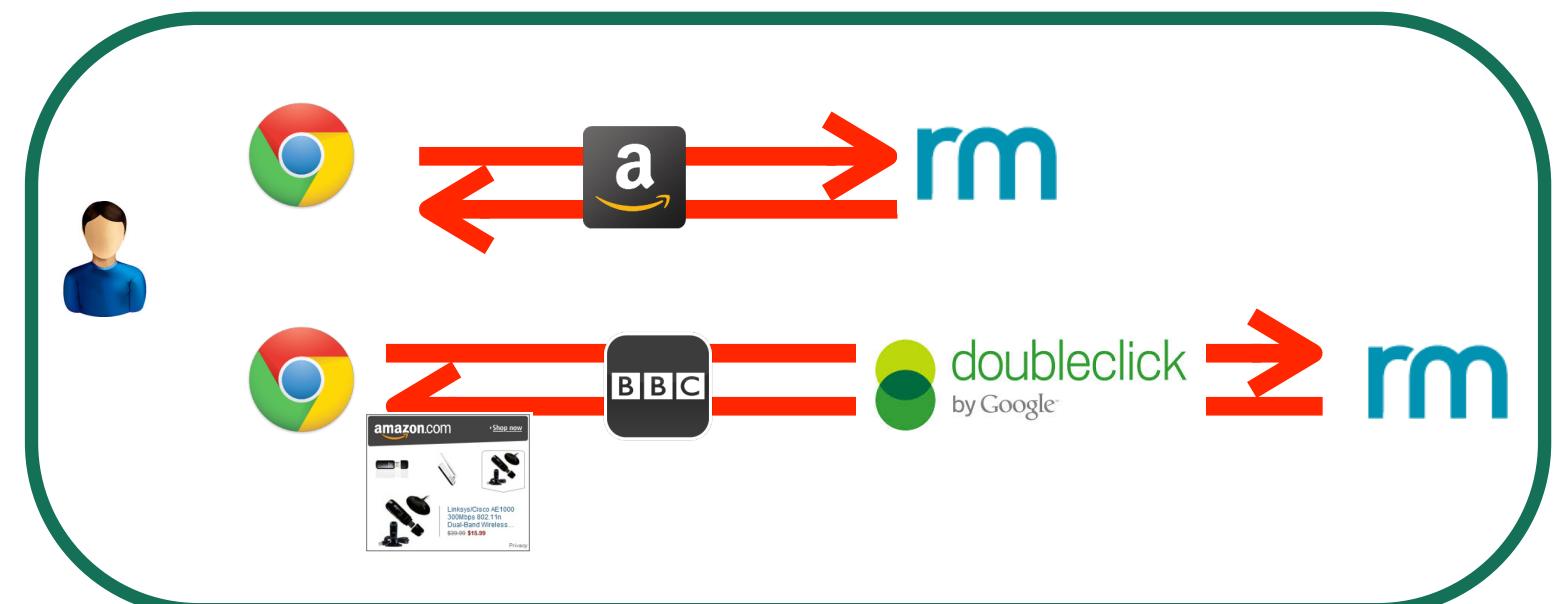
Now imagine that we have 100 users.



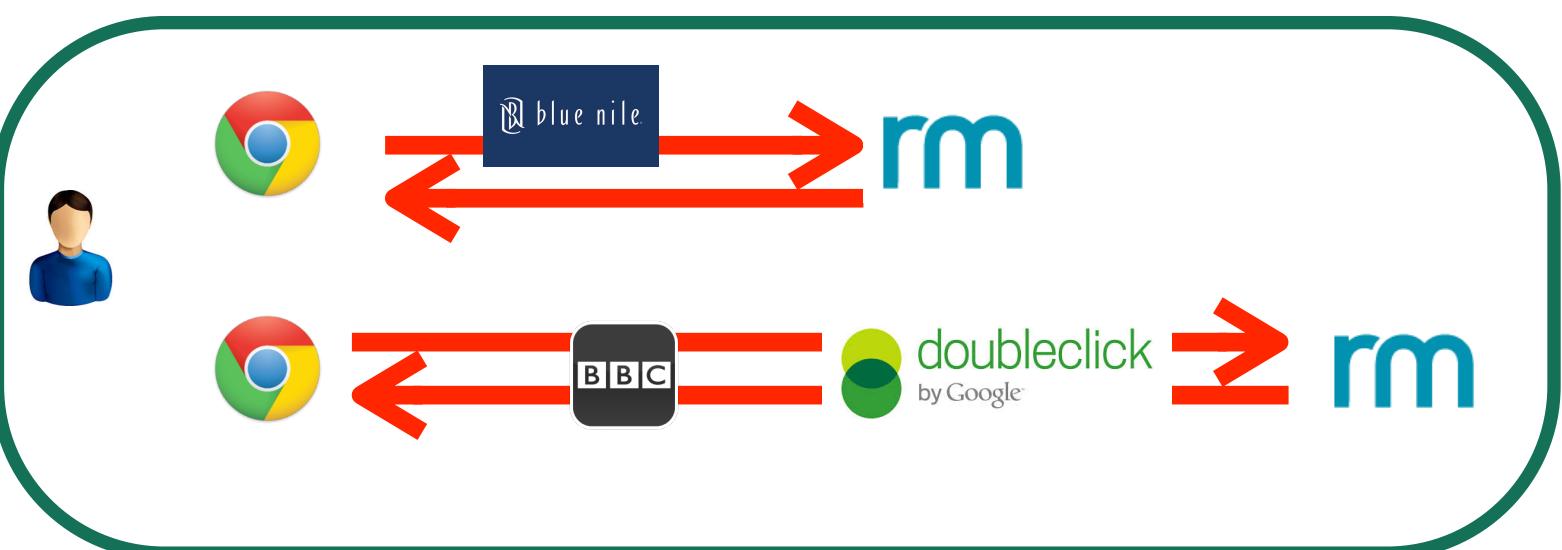
.



Controlled Experiments

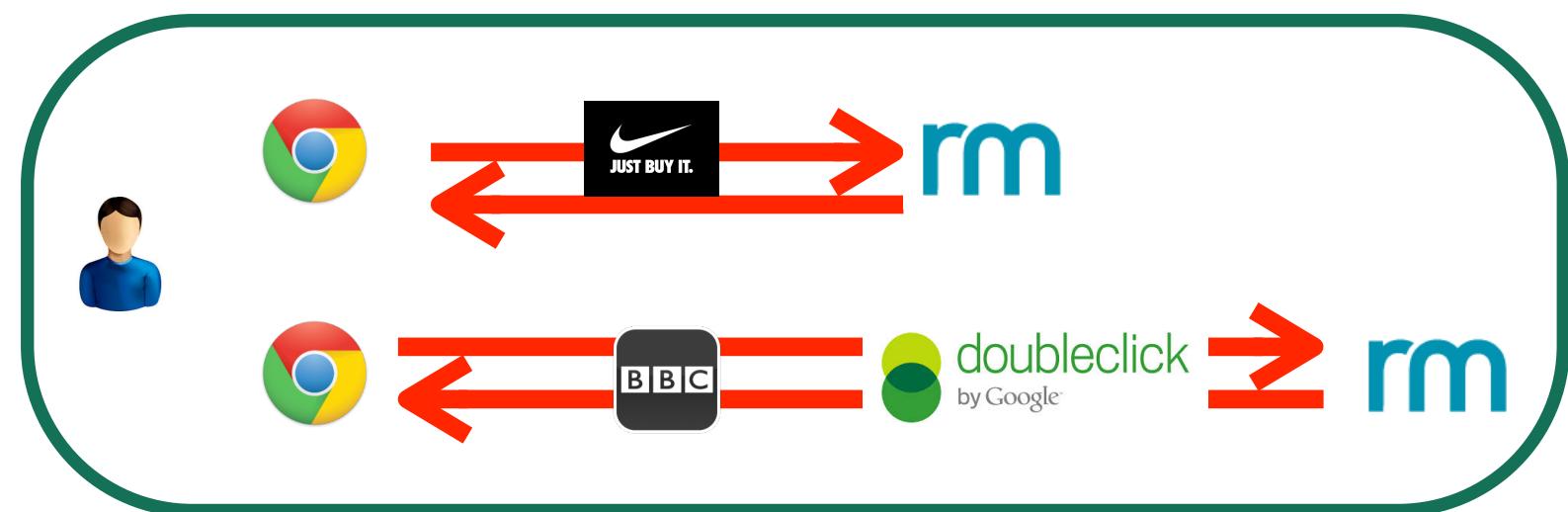


Now imagine that we have 100 users.



- Only the 1st user visits amazon.
- ... and eventually sees an ad for amazon product.

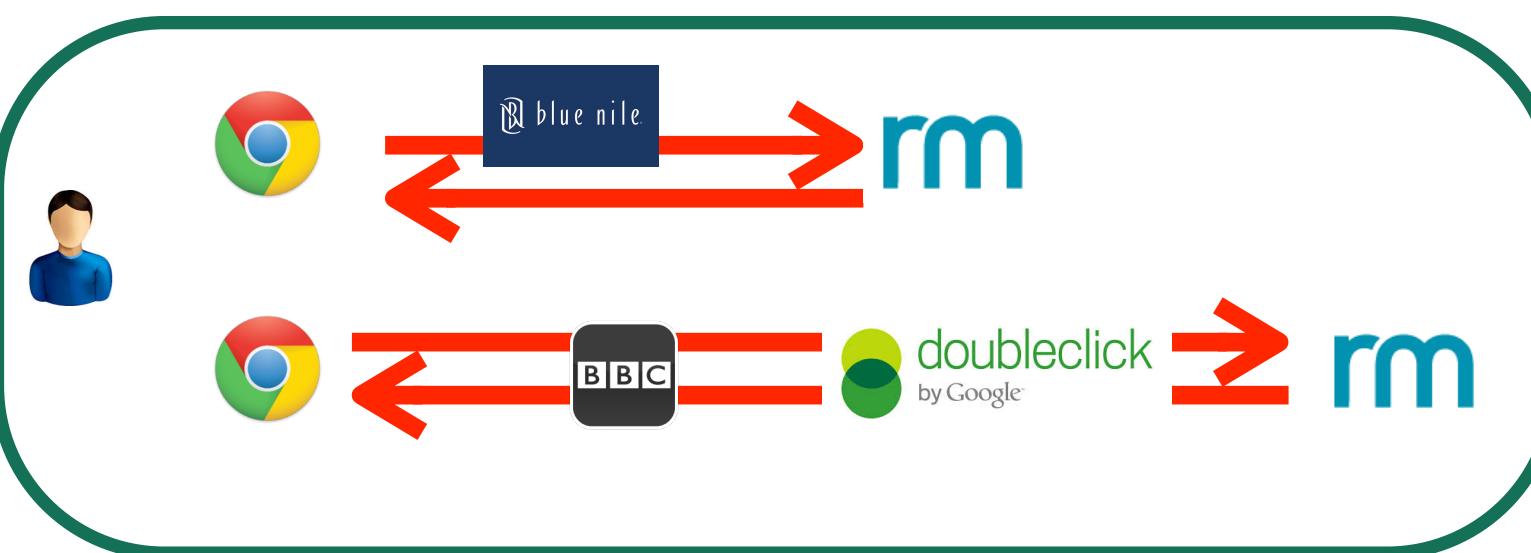
.



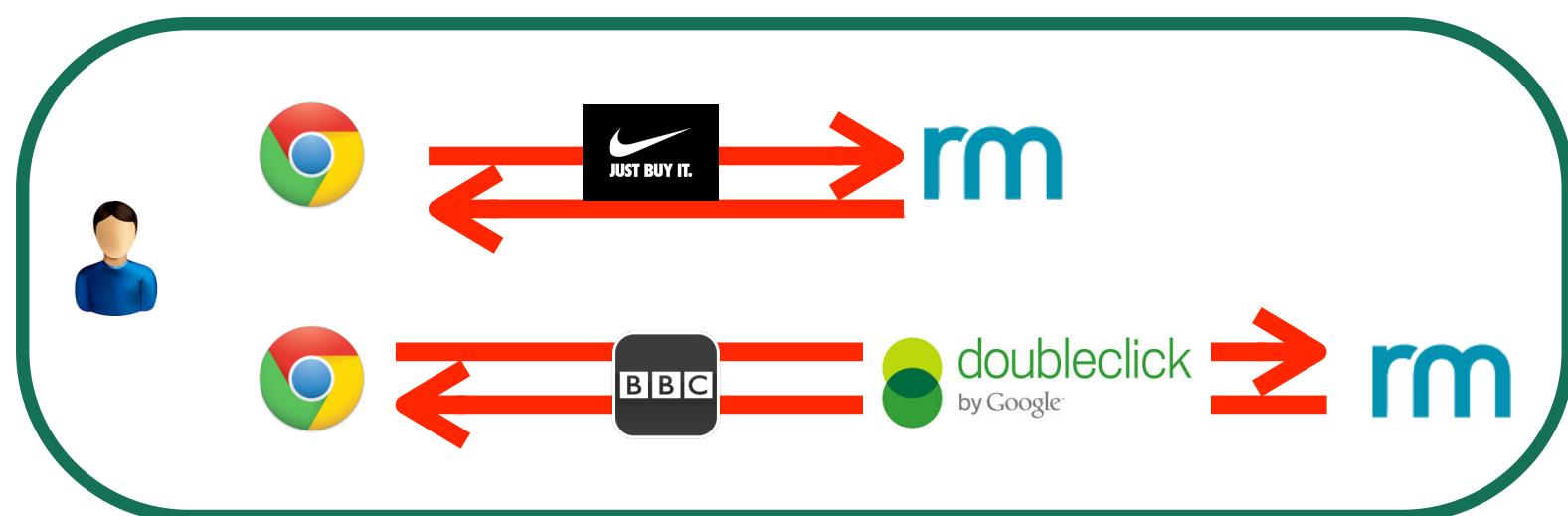
Controlled Experiments



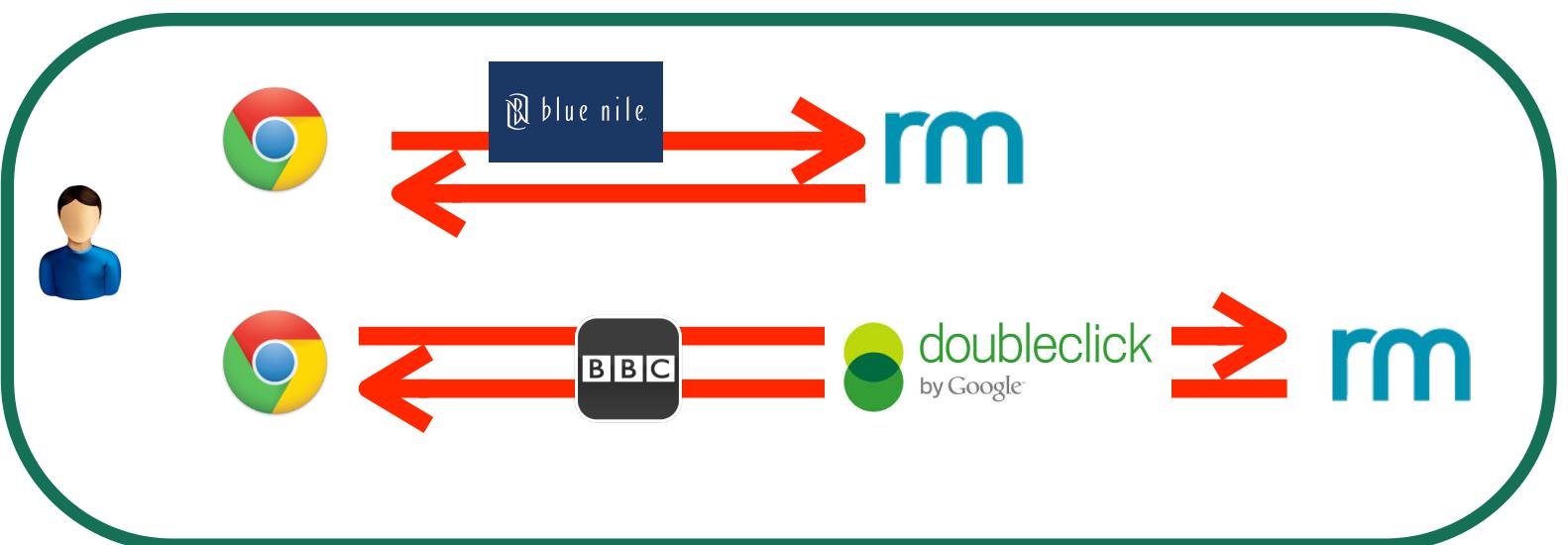
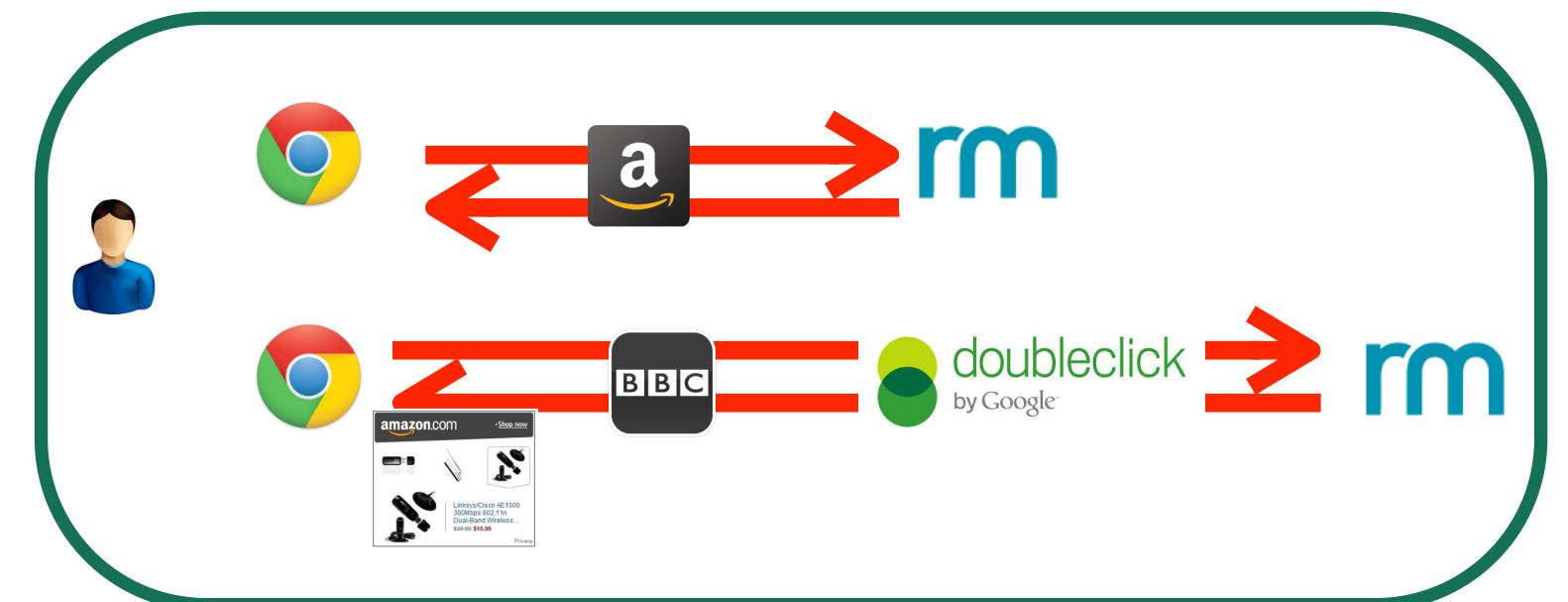
Now imagine that we have 100 users.



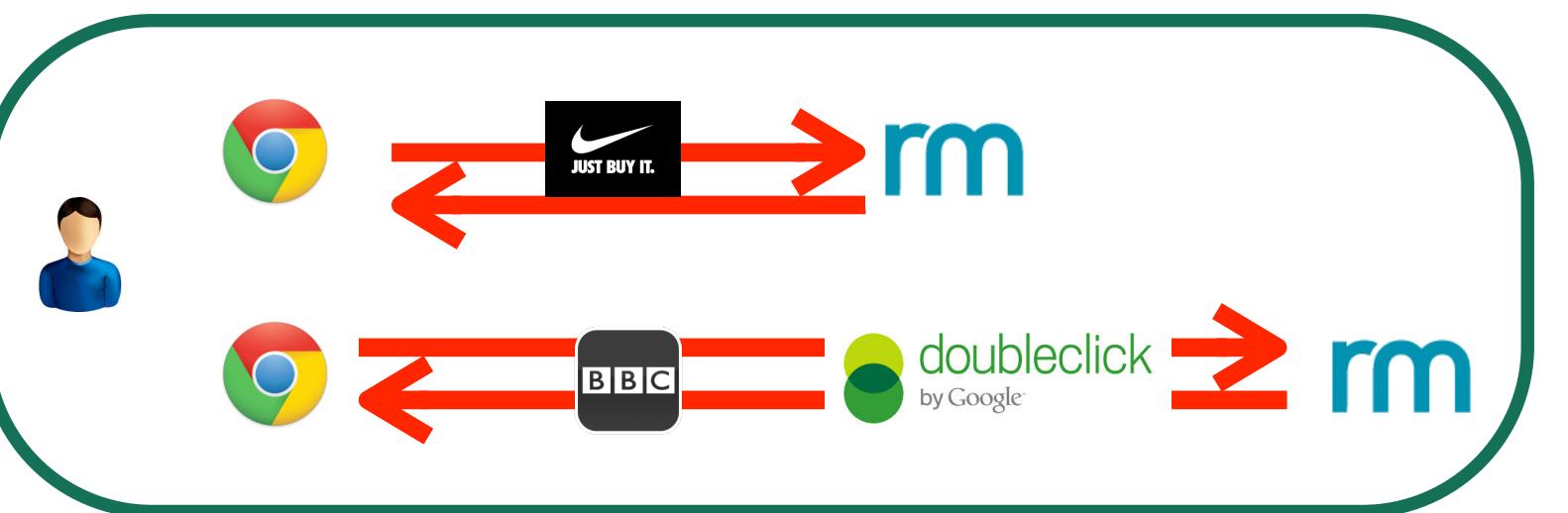
- Only the 1st user visits amazon.
- ... and eventually sees an ad for amazon product.
- Now we are much more confident about information sharing between DoubleClick and RightMedia.



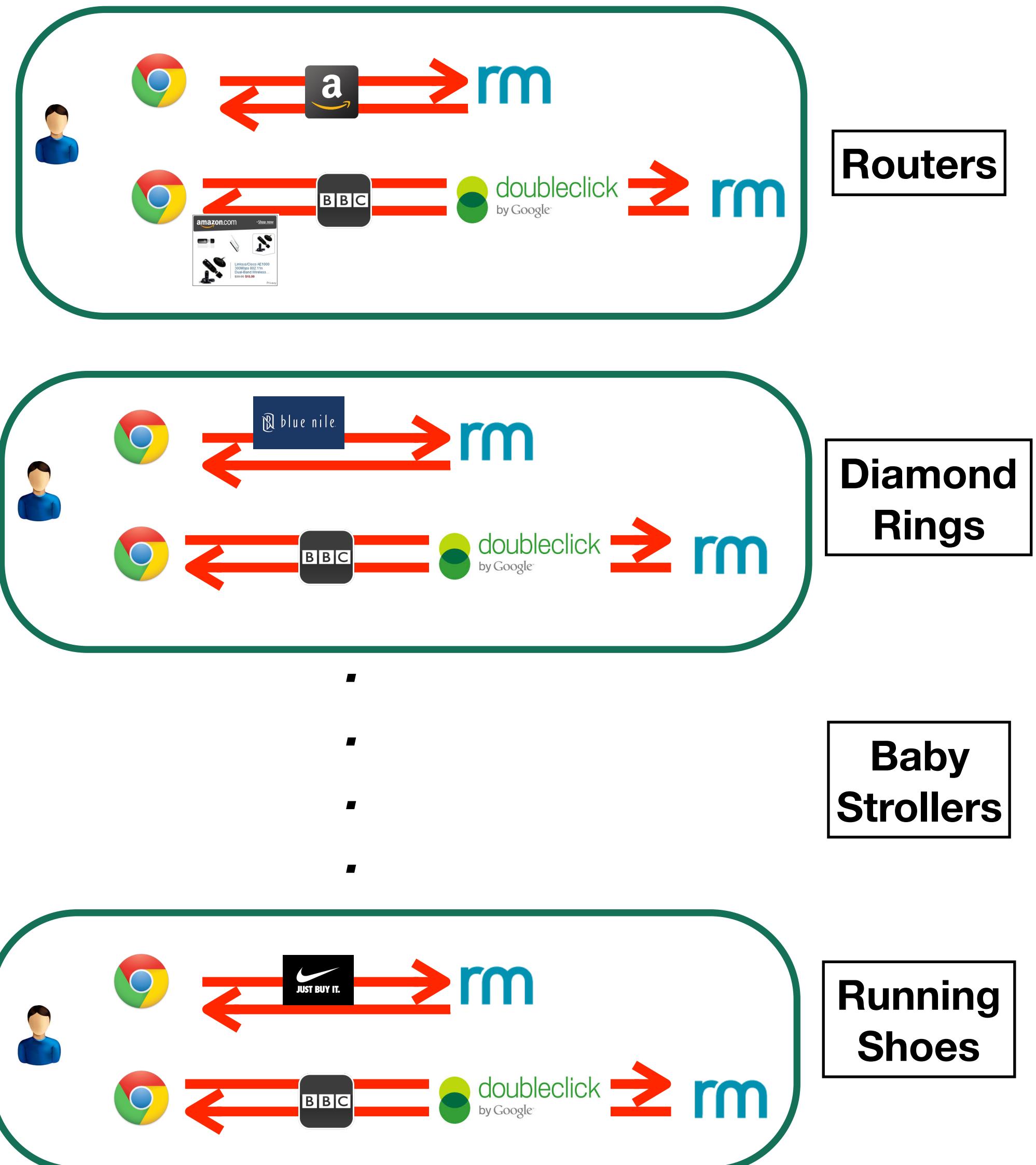
Controlled Experiments



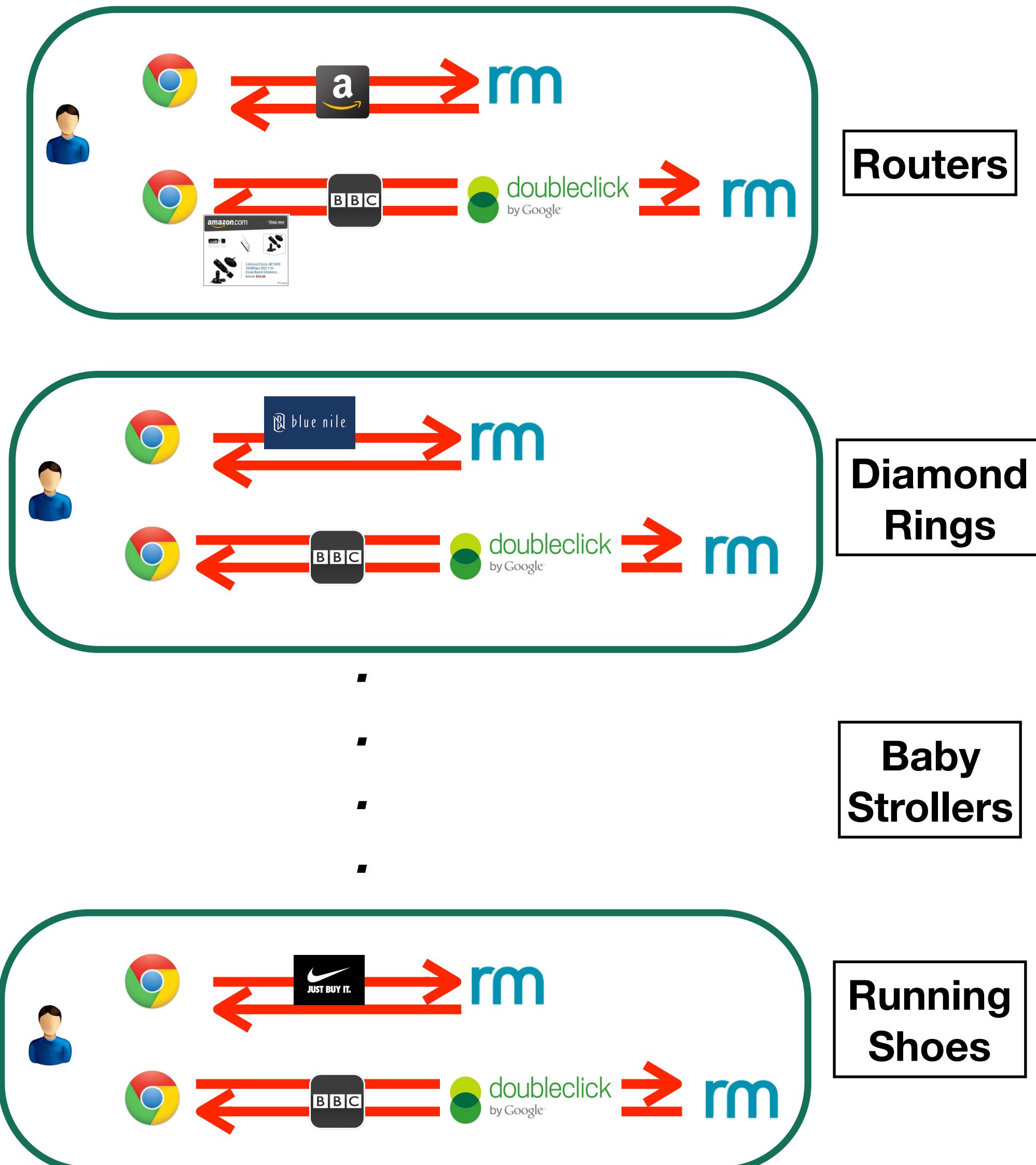
-



Controlled Experiments



Controlled Experiments



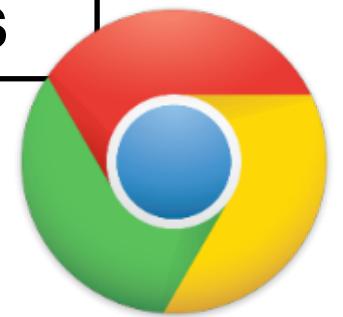
We set up 90 personas (interest profiles)

- Each Persona is very specific
- We select 10 websites per persona
- ... and 10 products per website

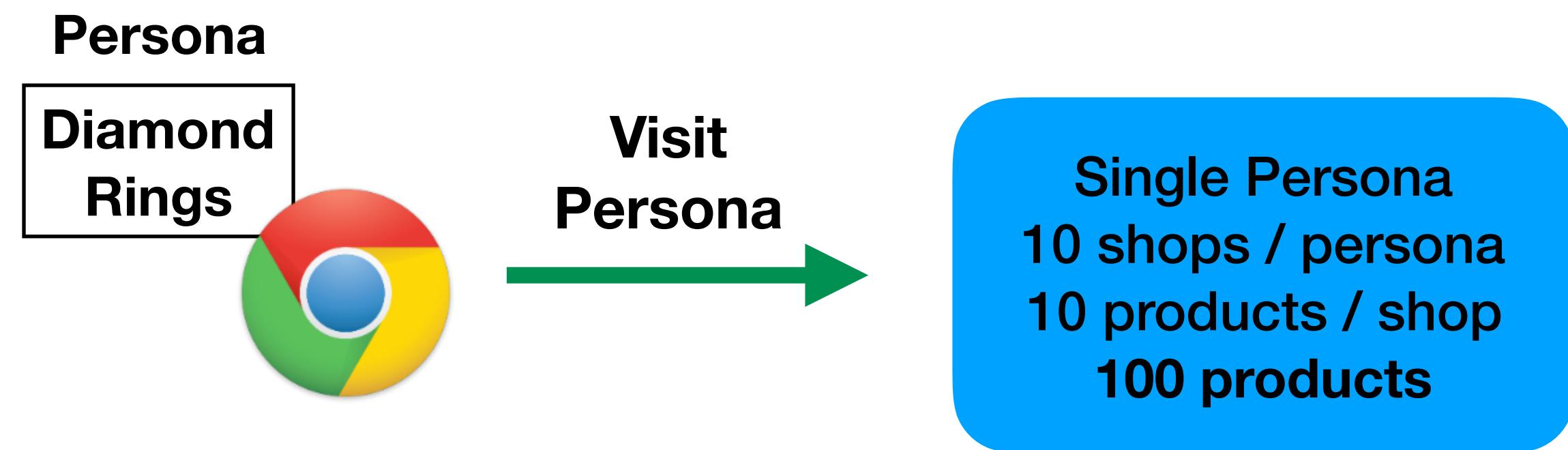
Experiment Setup & Data Collection

Persona

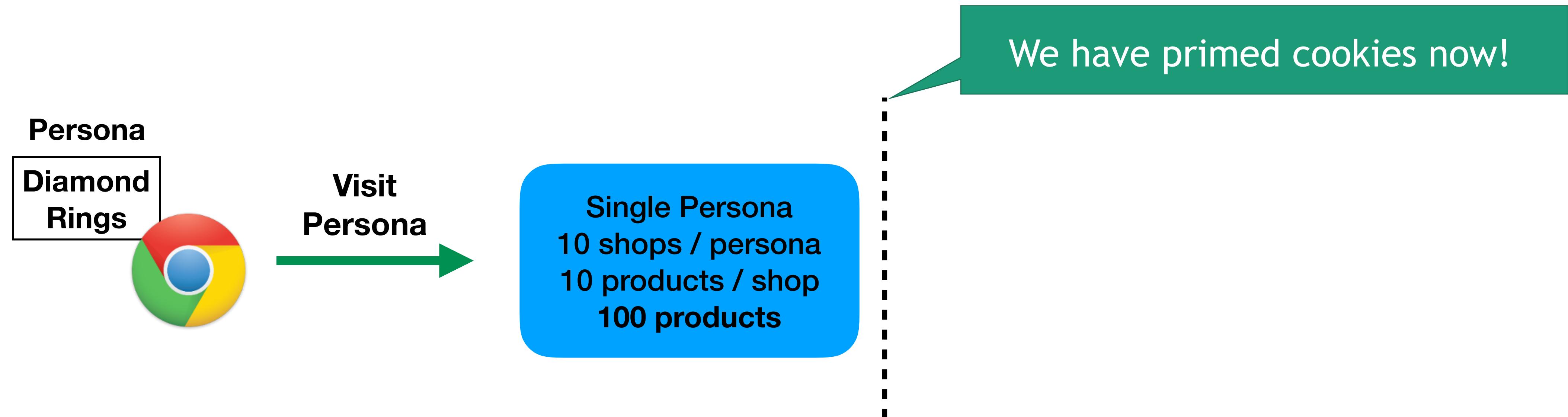
**Diamond
Rings**



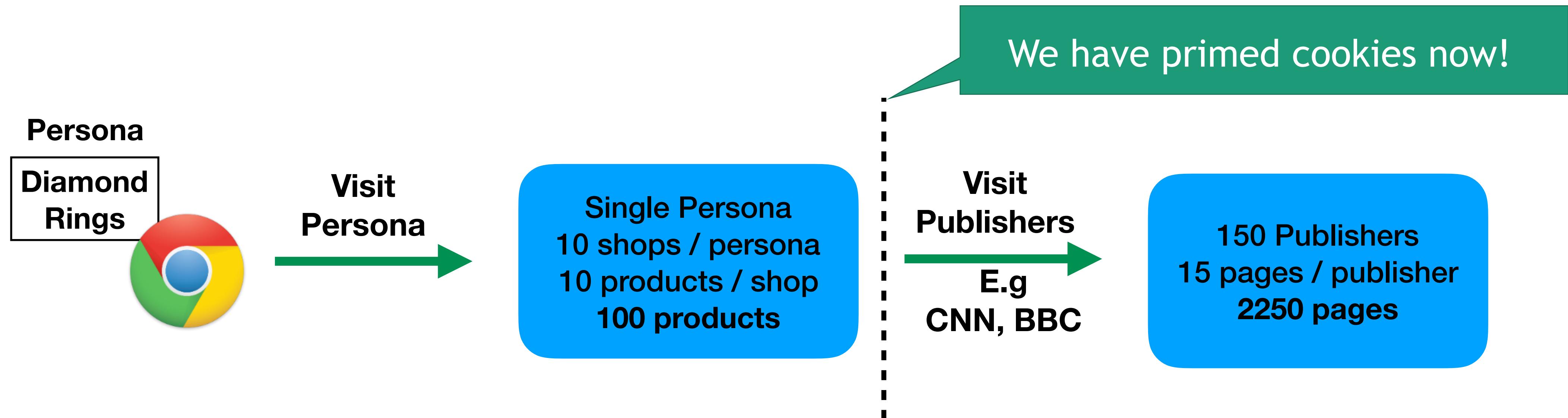
Experiment Setup & Data Collection



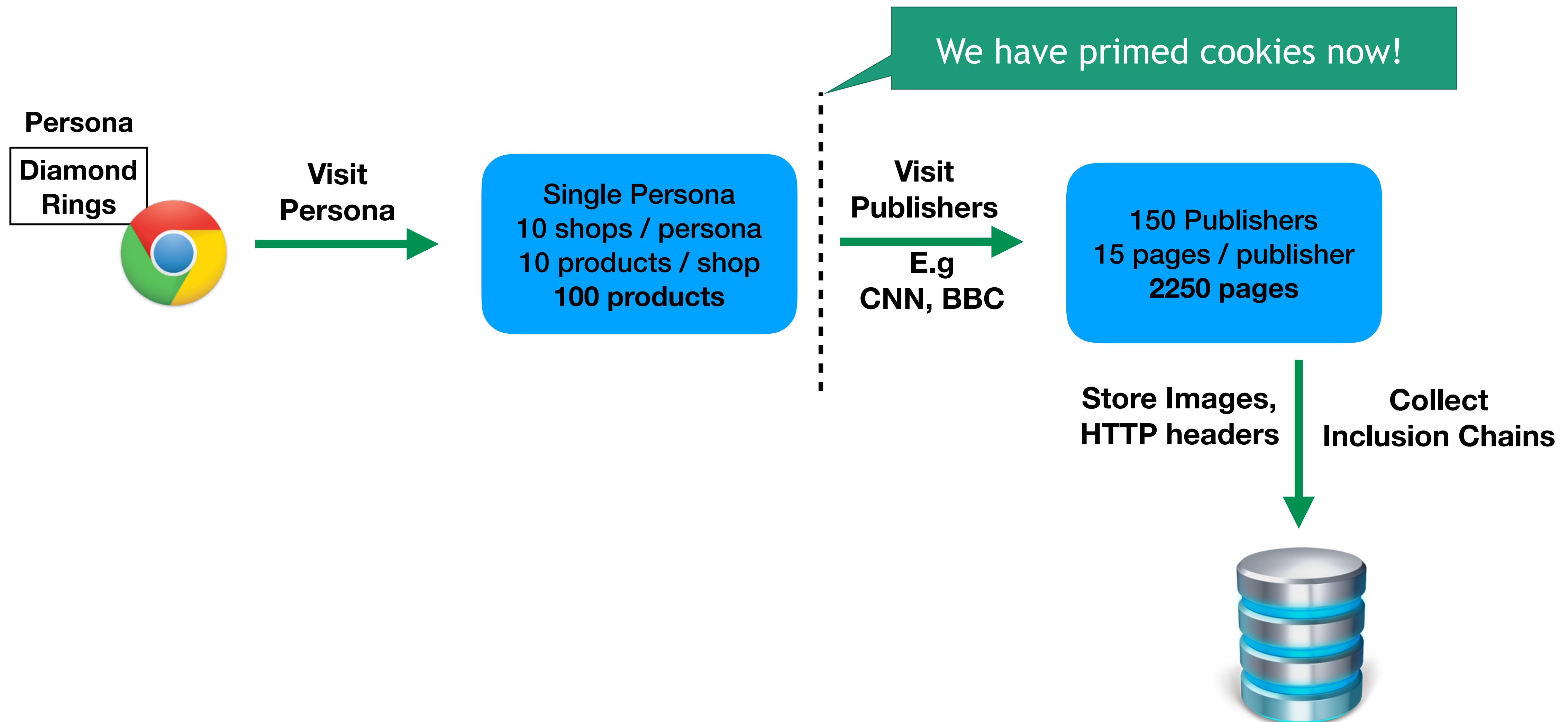
Experiment Setup & Data Collection



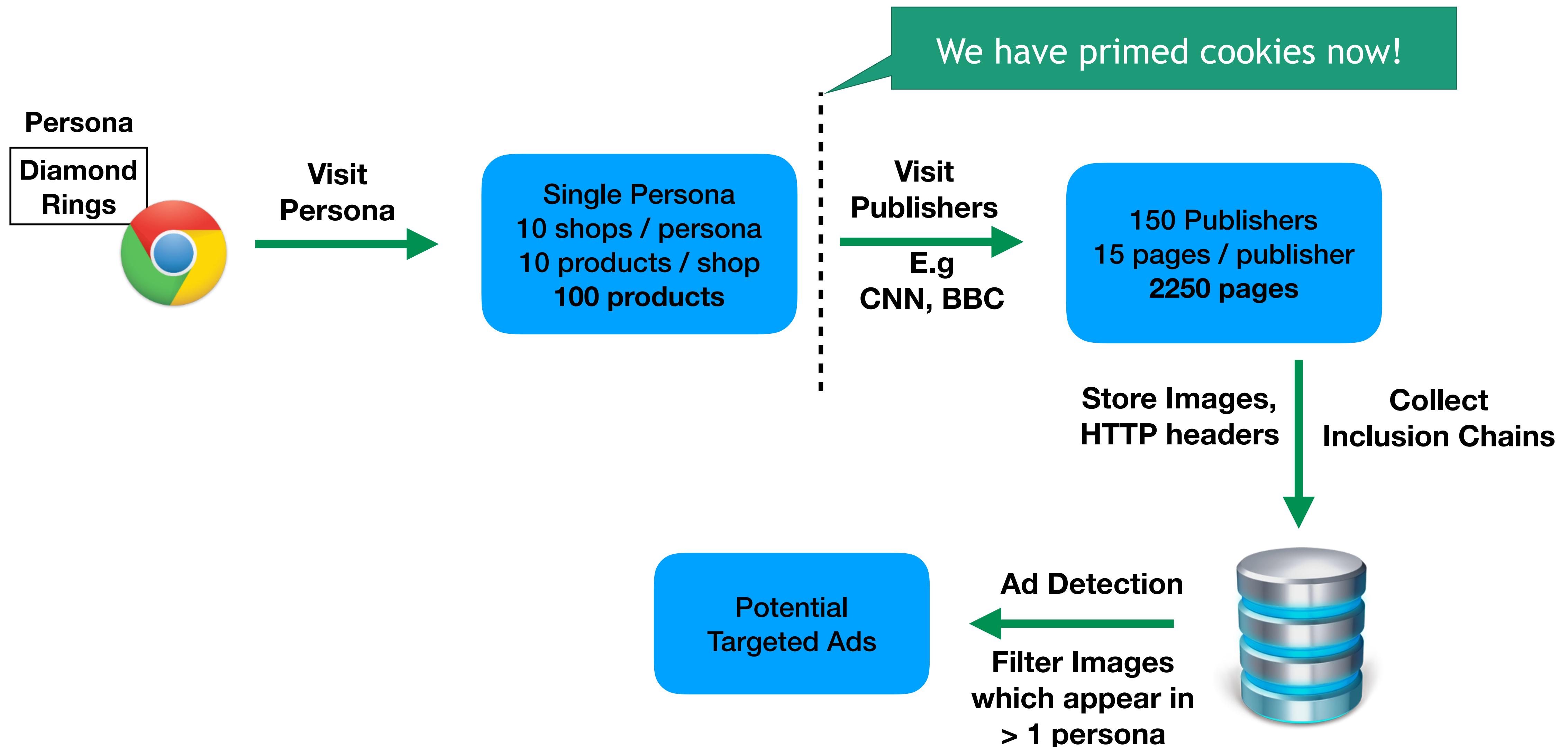
Experiment Setup & Data Collection



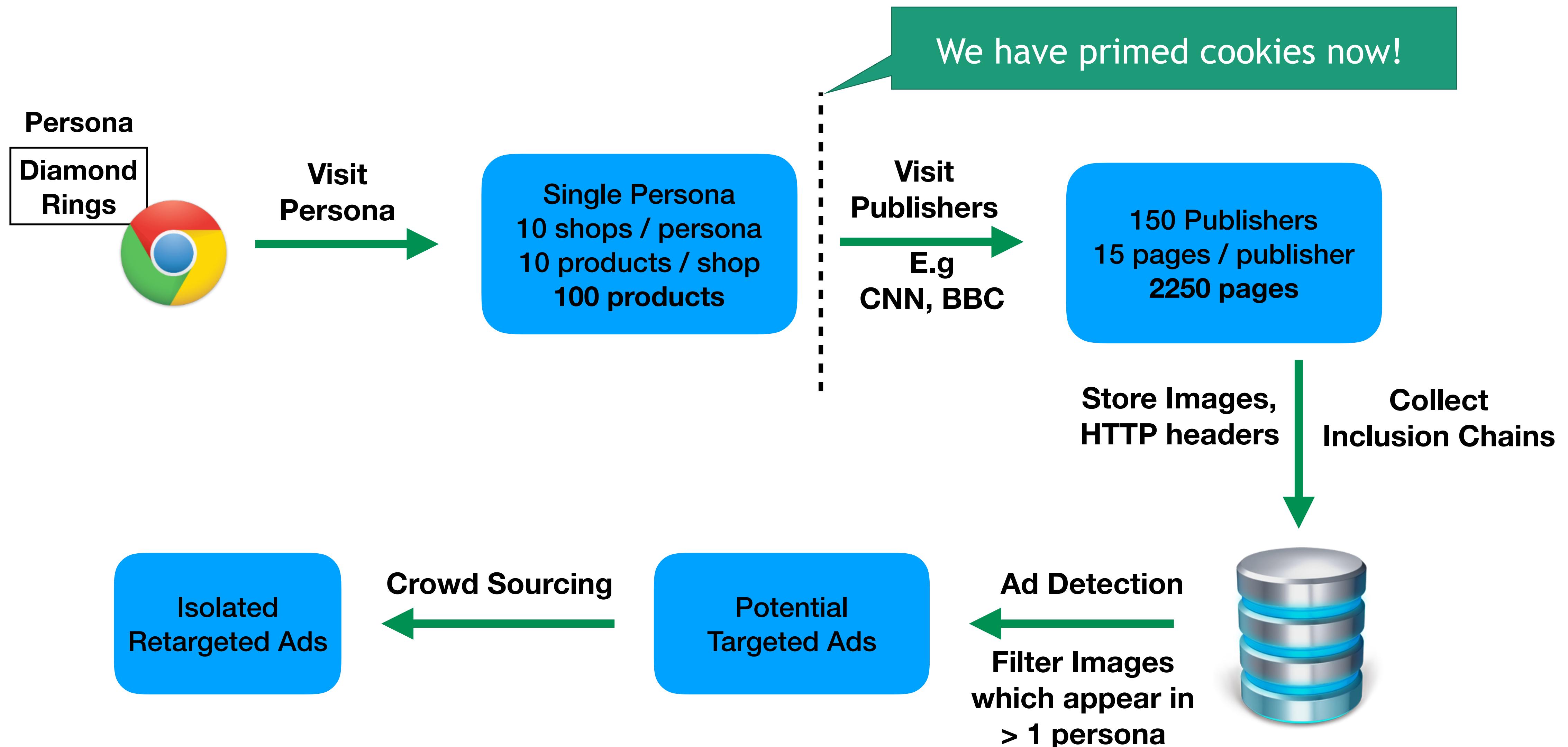
Experiment Setup & Data Collection



Experiment Setup & Data Collection



Experiment Setup & Data Collection



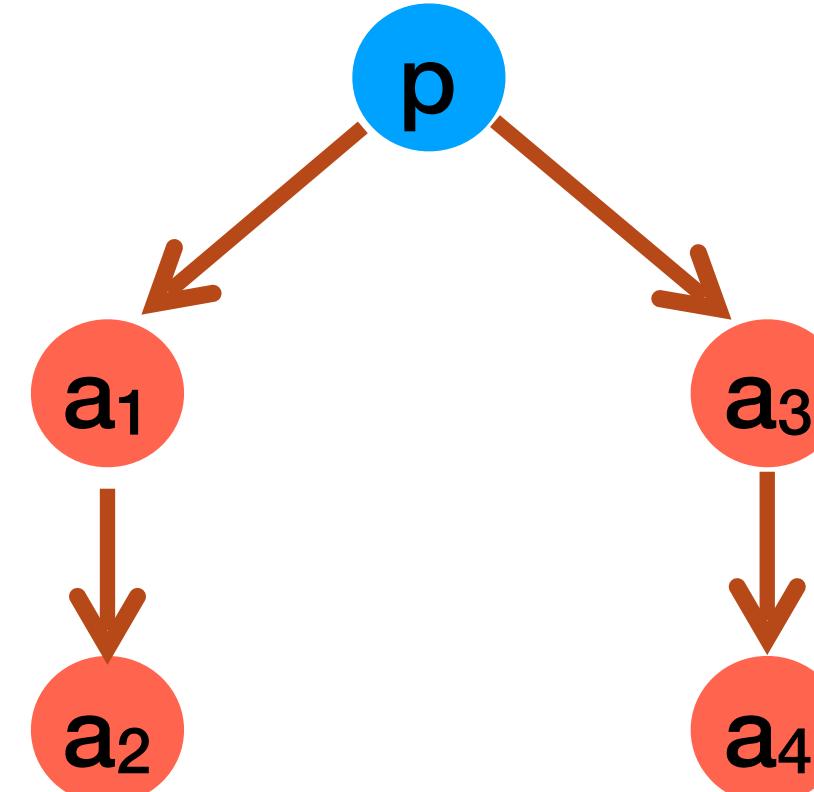
Inclusion Chains Associated With Retargeted Ads

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Inclusion Chains



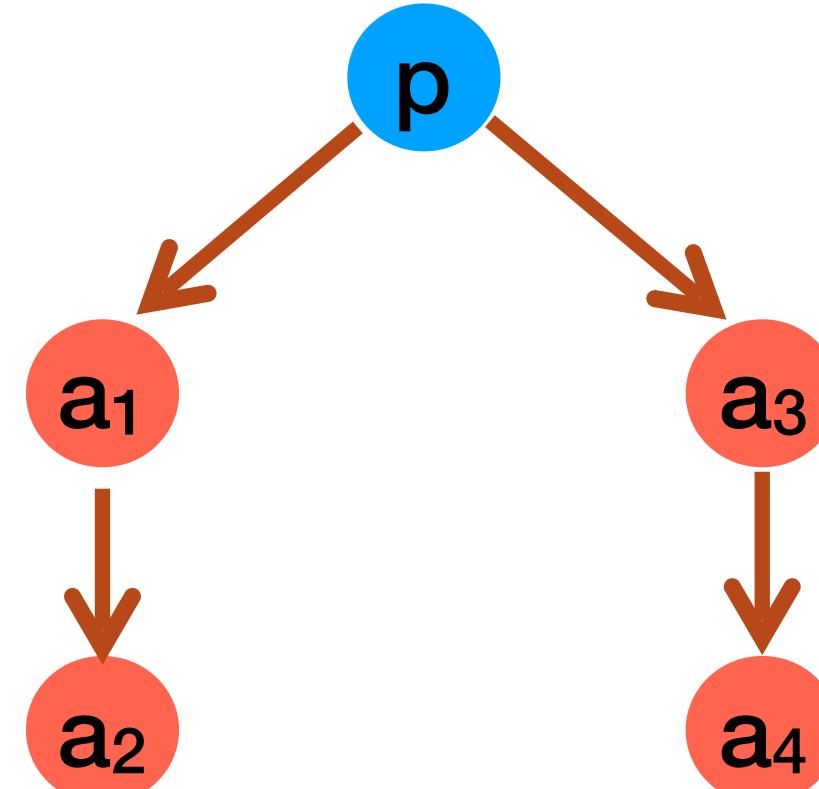
Inclusion Chains Associated With Retargeted Ads

DOM Tree for <http://p.com/index.html>

```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Inclusion Chains



- 5,102 unique retargeted ads
- 35,448 associated **publisher-side** chains
 - We observed some retargets multiple times

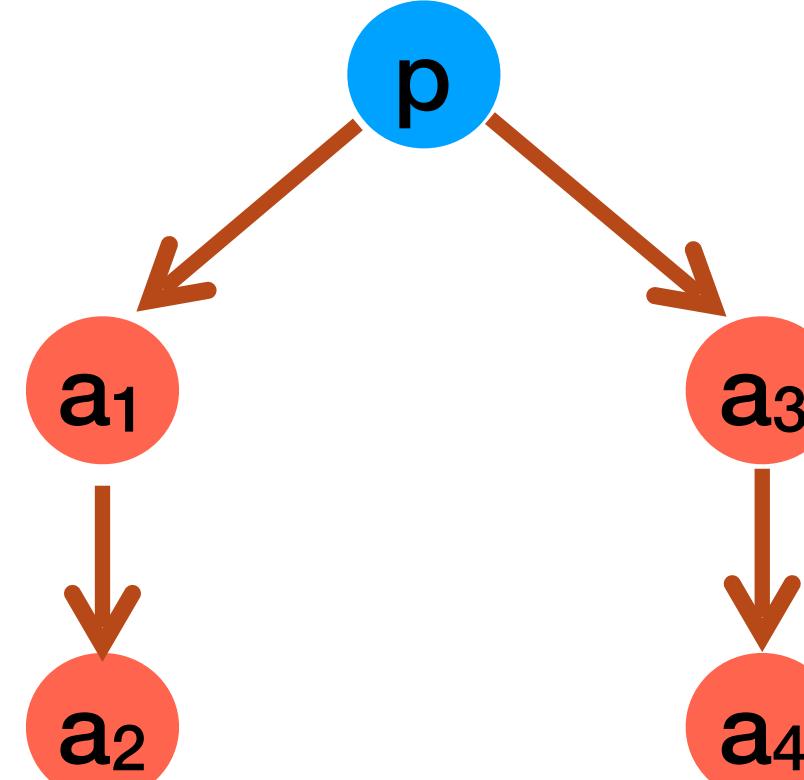
Inclusion Chains Associated With Retargeted Ads

DOM Tree for <http://p.com/index.html>

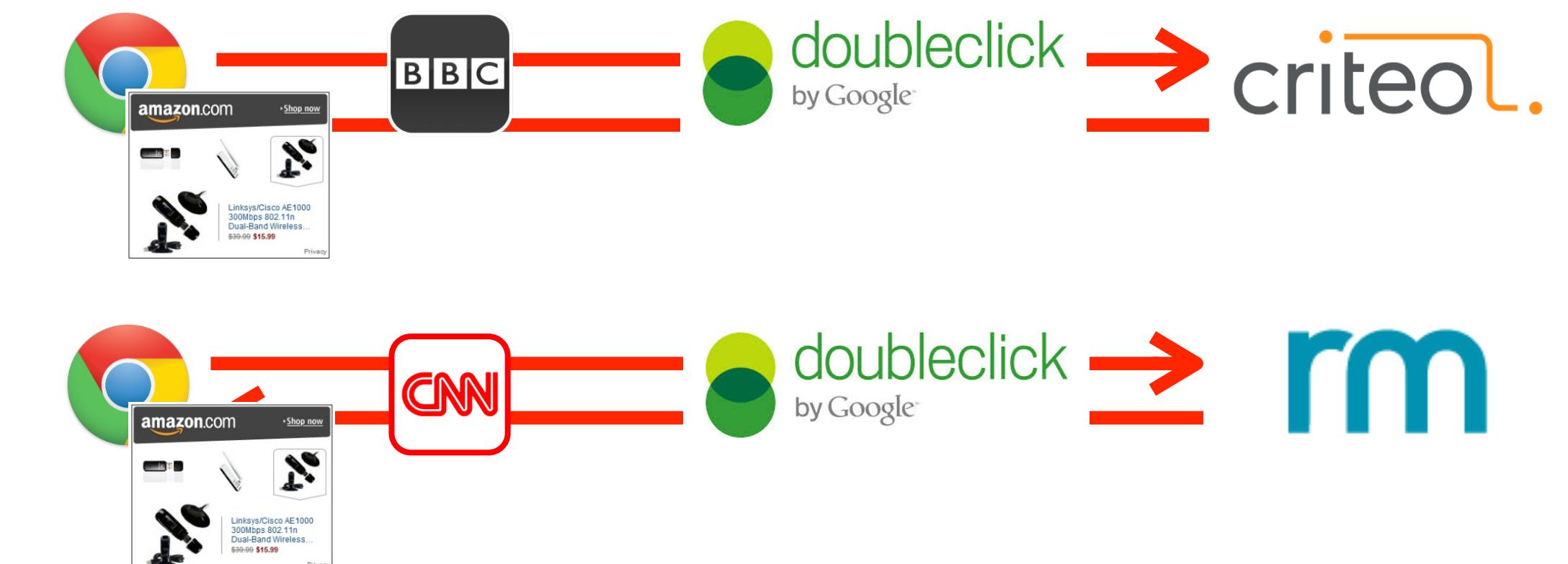
```
<html>
  <body>
    <script src="a1.com/cookie-match.js"></script>
    <!-- Tracking pixel inserted dynamically
        by cookie-match.js -->
    

    <iframe src="a3.com/banner.html">
      <script src="a4.com/ads.js"></script>
    </iframe>
  </body>
</html>
```

Inclusion Chains



- 5,102 unique retargeted ads
- 35,448 associated publisher-side chains
 - We observed some retargets multiple times



Results

Categorized 35K chains that serve retargeted ads into 4 categories

Results

Categorized 35K chains that serve retargeted ads into 4 categories

- Cookie Matching was the most common matching type – 71%

Results

Categorized 35K chains that serve retargeted ads into 4 categories

- Cookie Matching was the most common matching type – 71%
- 15% of the chains were categorized as server-side matching
 - All of these instances were Google domains sharing data on server side

Results

Categorized 35K chains that serve retargeted ads into 4 categories

- Cookie Matching was the most common matching type – 71%
- 15% of the chains were categorized as server-side matching
 - All of these instances were Google domains sharing data on server side

Identified 200 cookie matching partner pairs

- 31% would have been missed by prior techniques

Results

Categorized 35K chains that serve retargeted ads into 4 categories

- Cookie Matching was the most common matching type – 71%
- 15% of the chains were categorized as server-side matching
 - All of these instances were Google domains sharing data on server side

Identified 200 cookie matching partner pairs

- 31% would have been missed by prior techniques

This methodology is content- and mechanism-agnostic

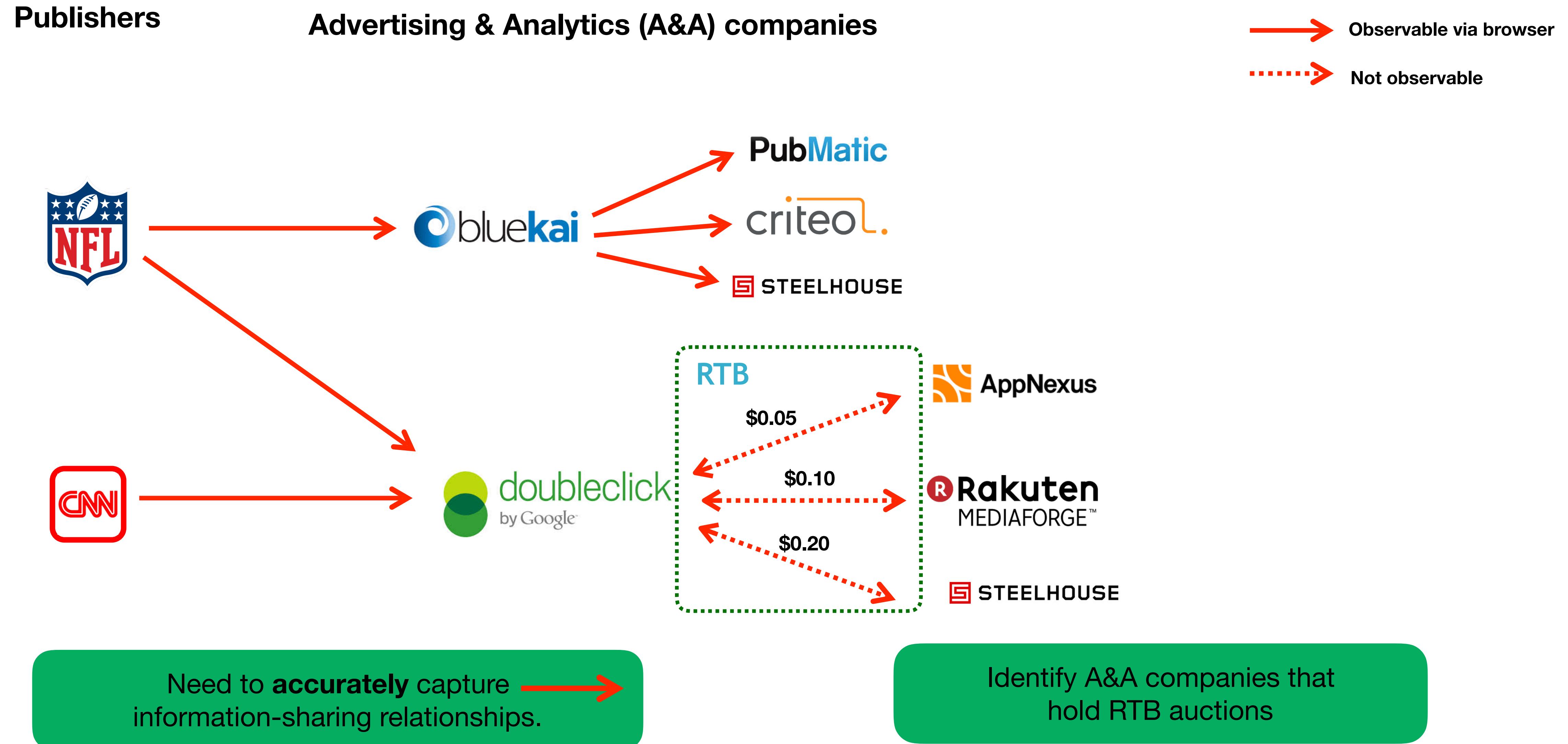
Thesis Outline

Detecting Information Sharing

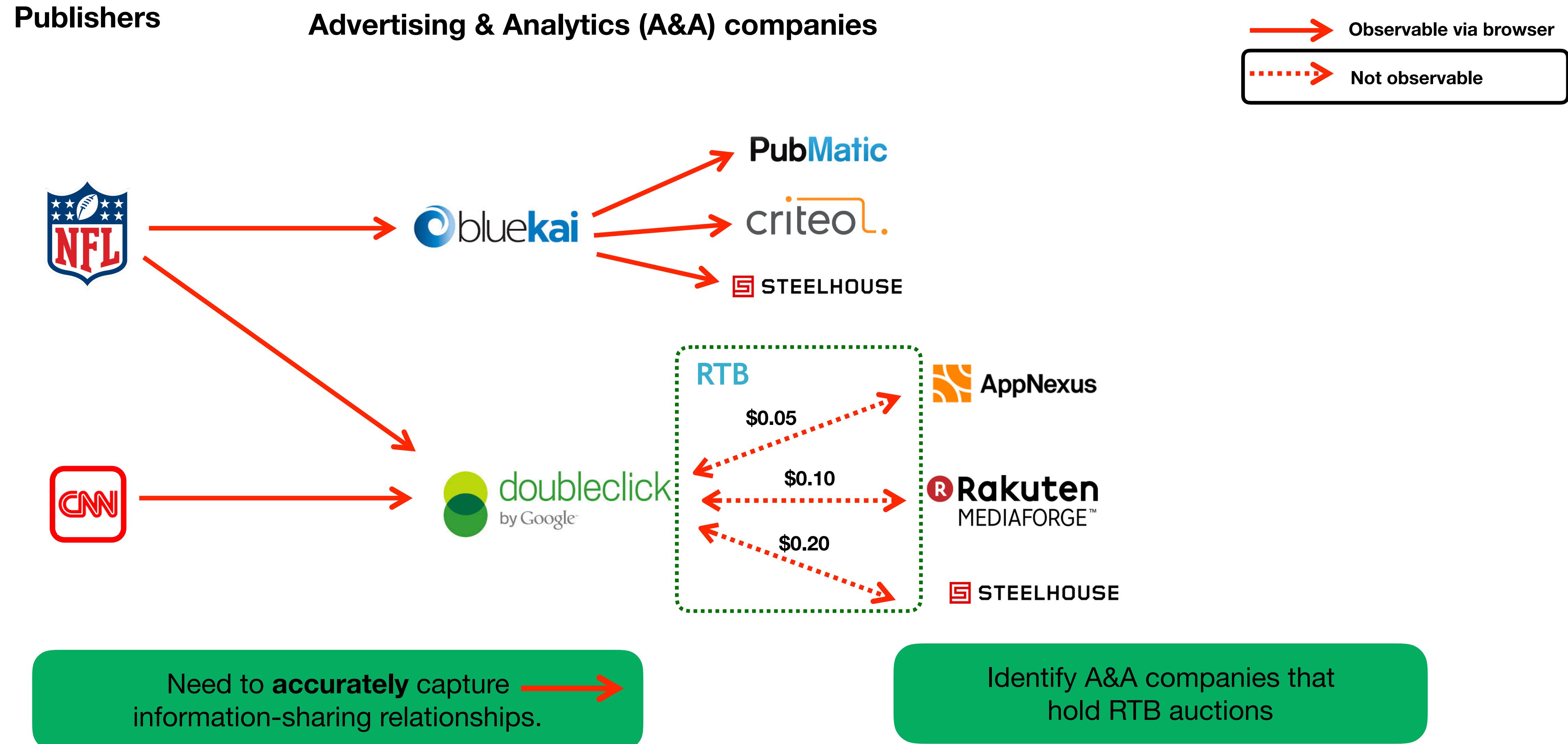
Isolating List of Ad Exchanges

Modeling User's Privacy Digital Footprint

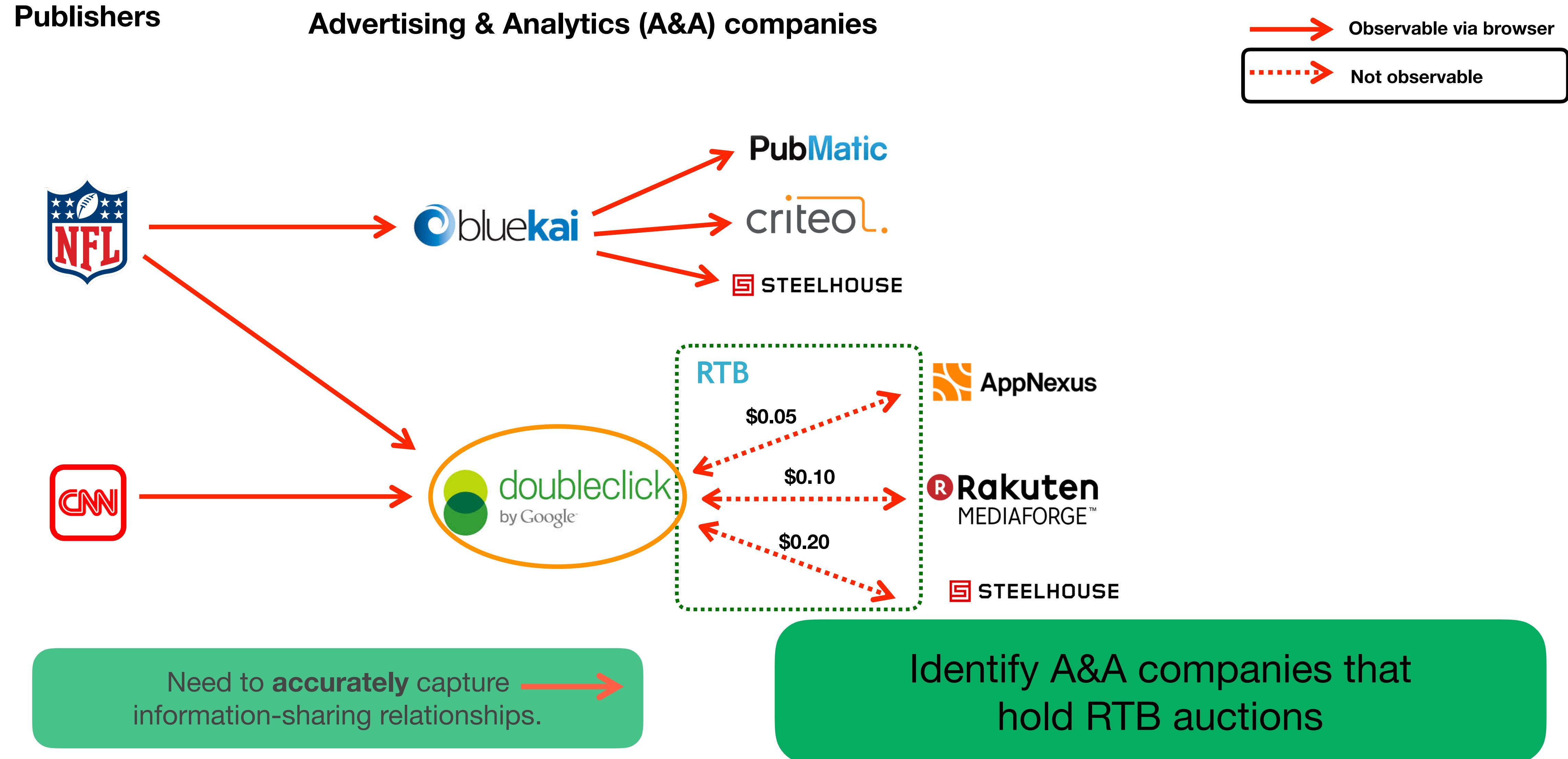
What Do We Need to Model Privacy Footprint?



What Do We Need to Model Privacy Footprint?



What Do We Need to Model Privacy Footprint?

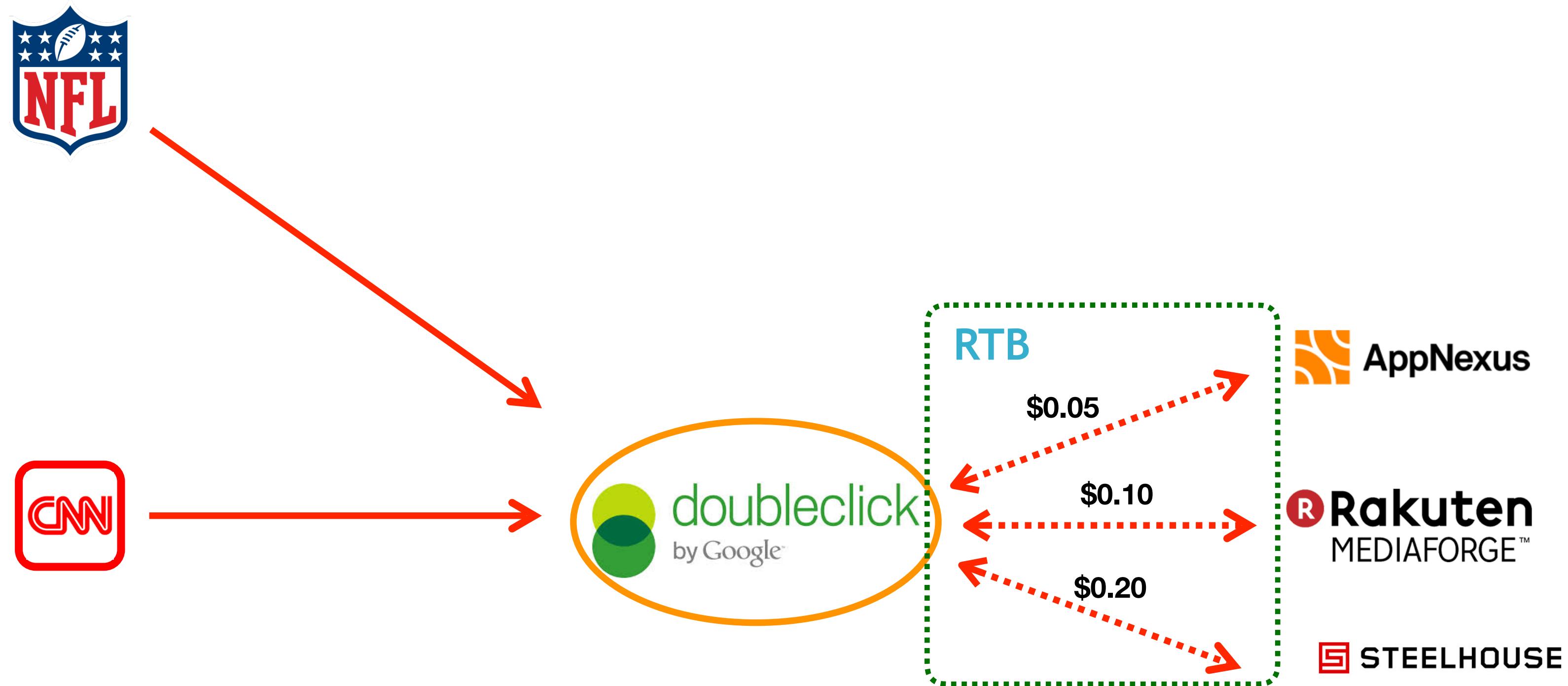


The ‘Power’ of Ad Exchanges

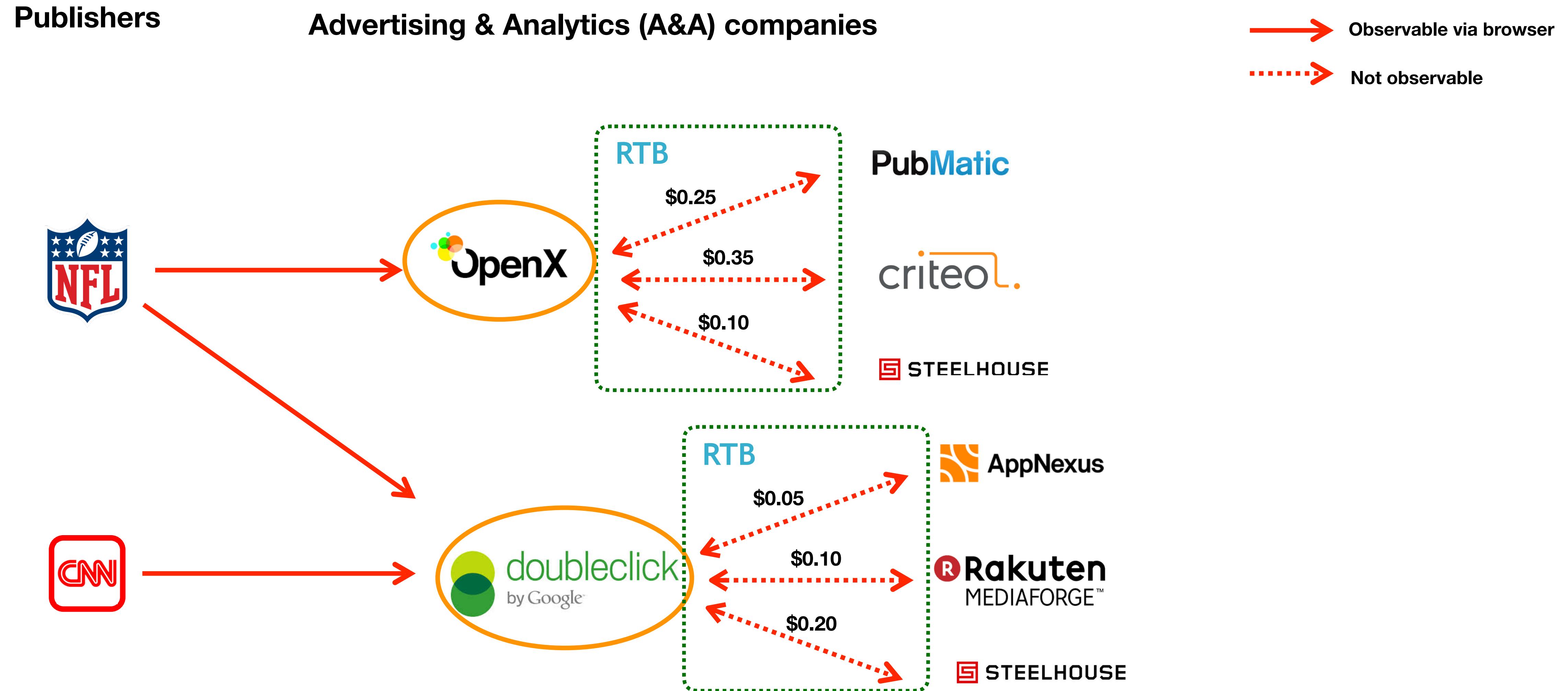
Publishers

Advertising & Analytics (A&A) companies

- Observable via browser
- Not observable



The ‘Power’ of Ad Exchanges

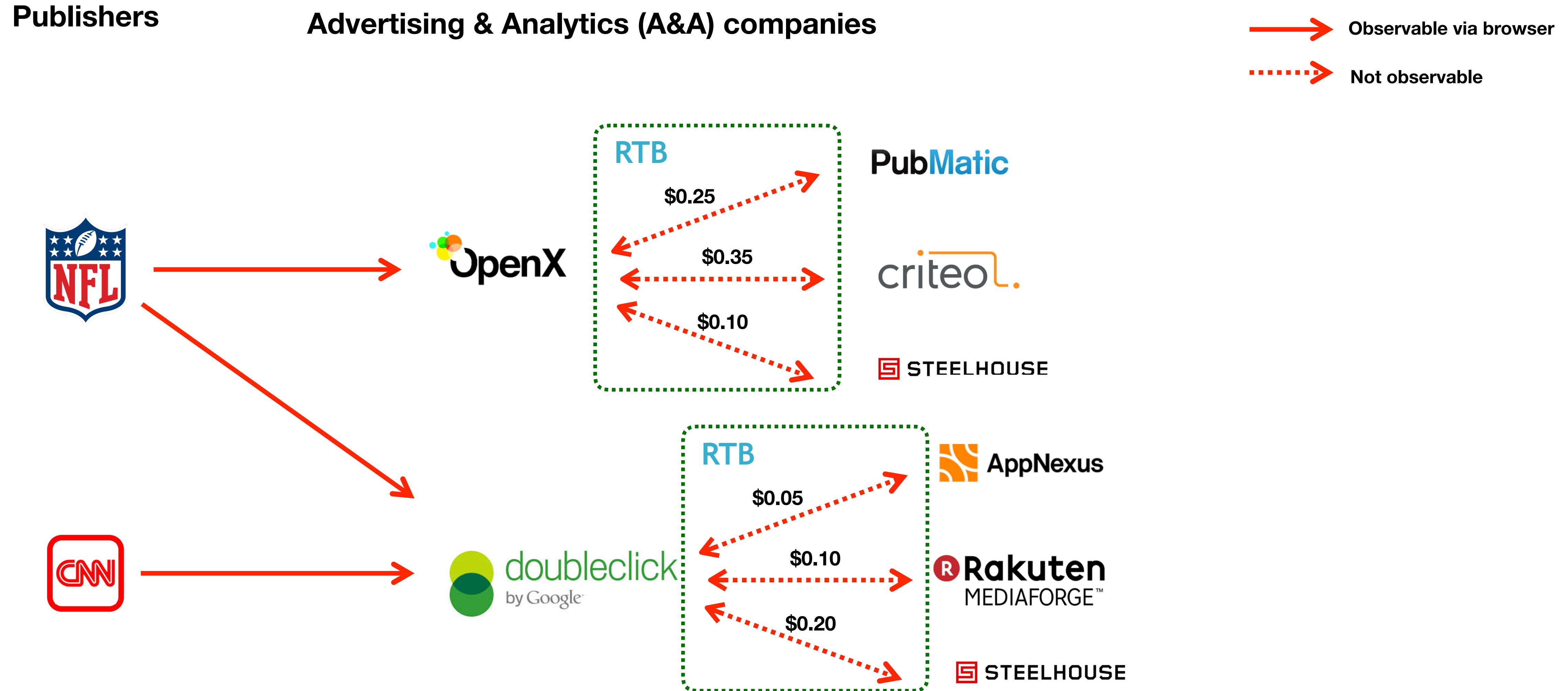


High-Level Idea

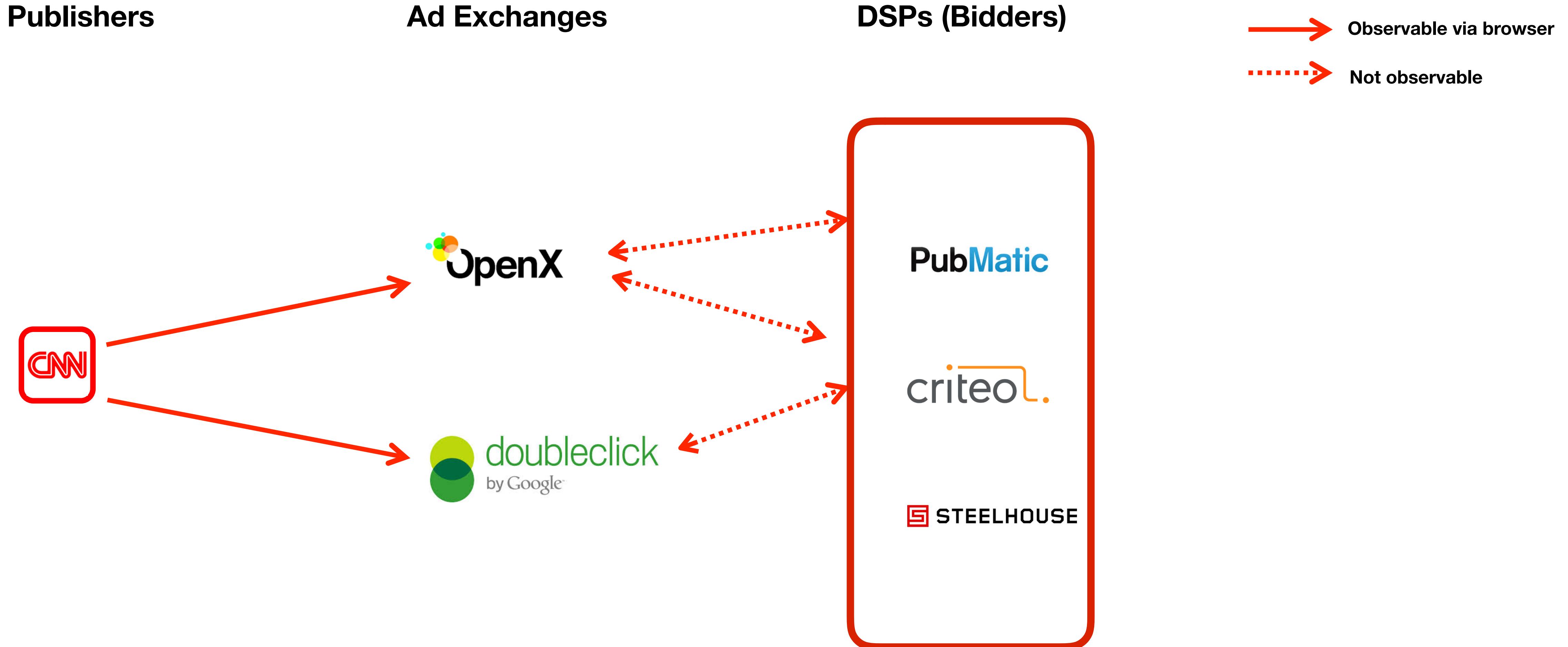
**Use ads.txt transparency standard to isolate
a list of Ad Exchanges**

Introduced to combat ‘domain spoofing’ fraud in RTB auctions

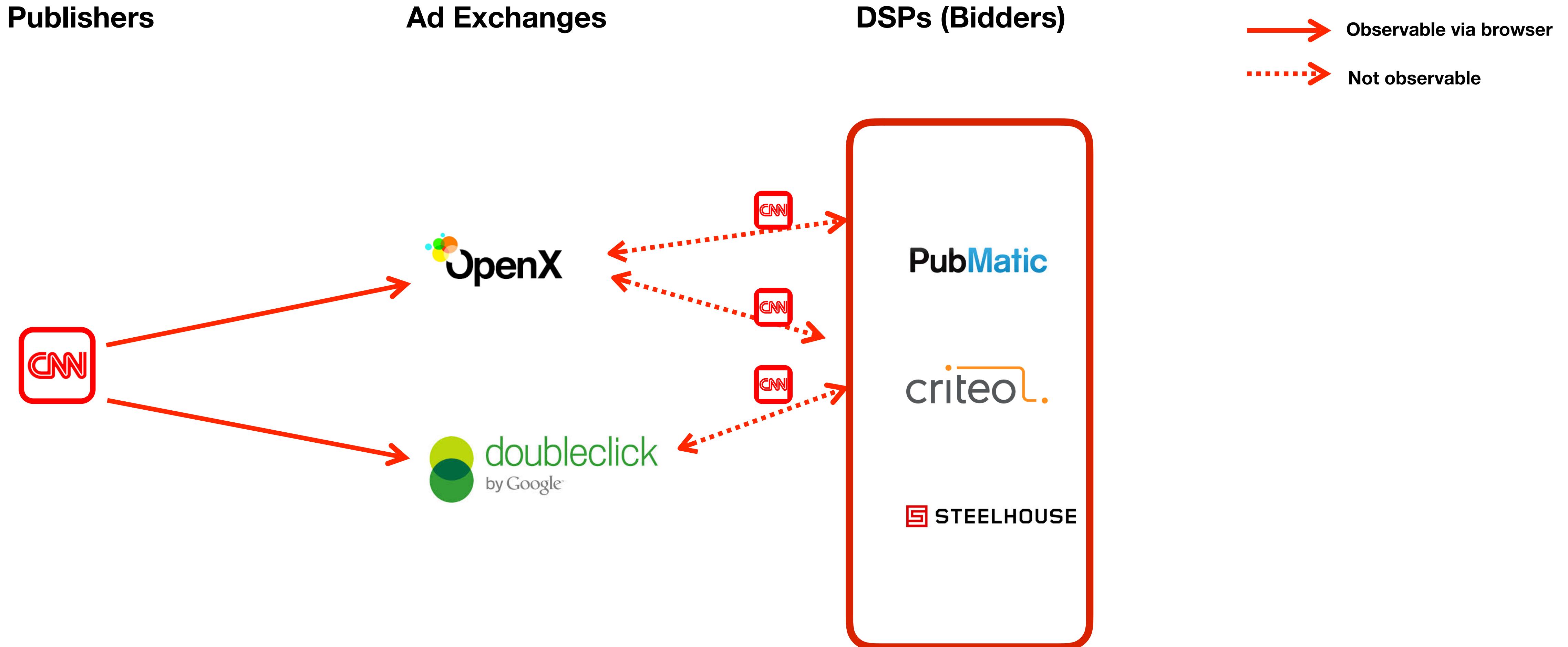
Potential for Fraud in RTB



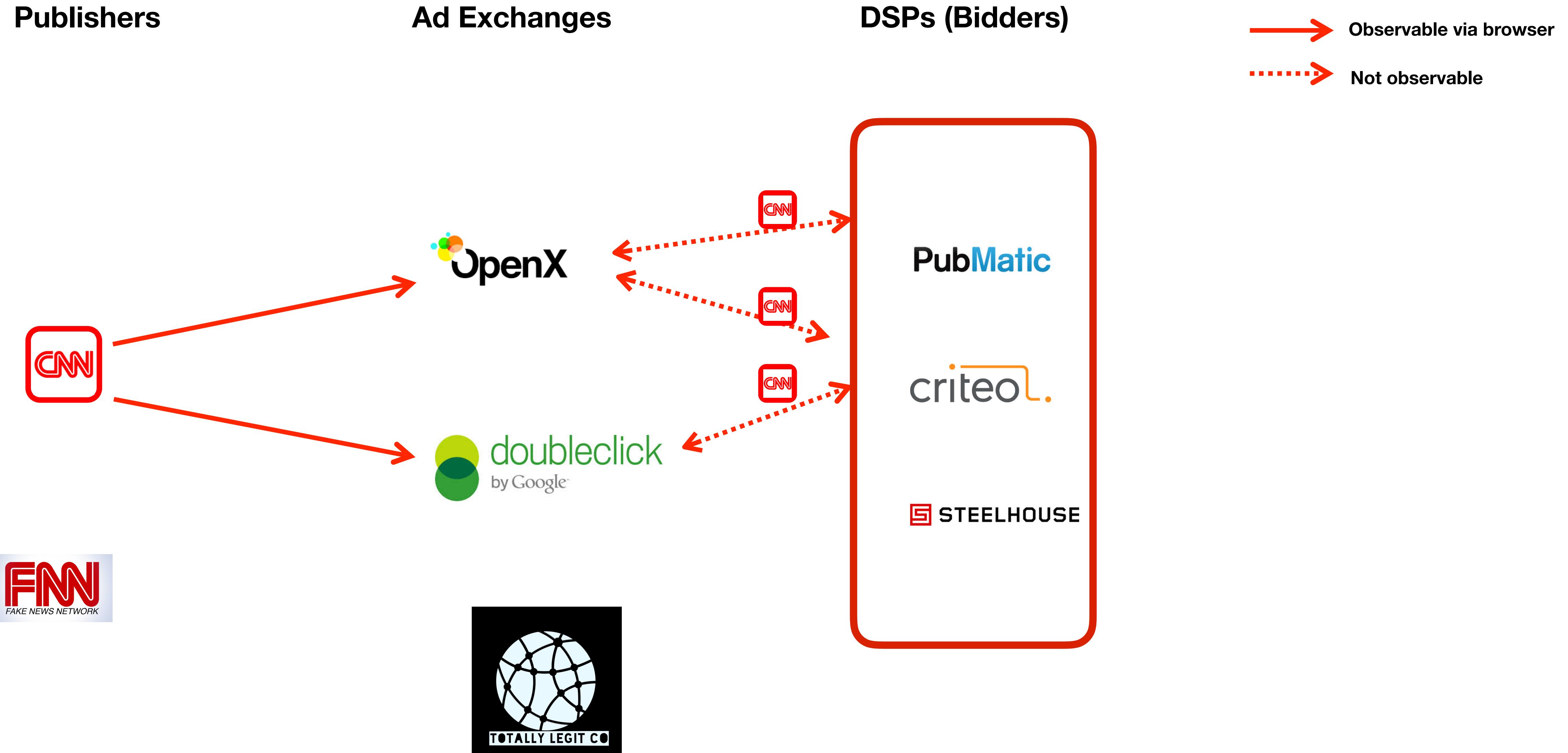
Potential for Fraud in RTB



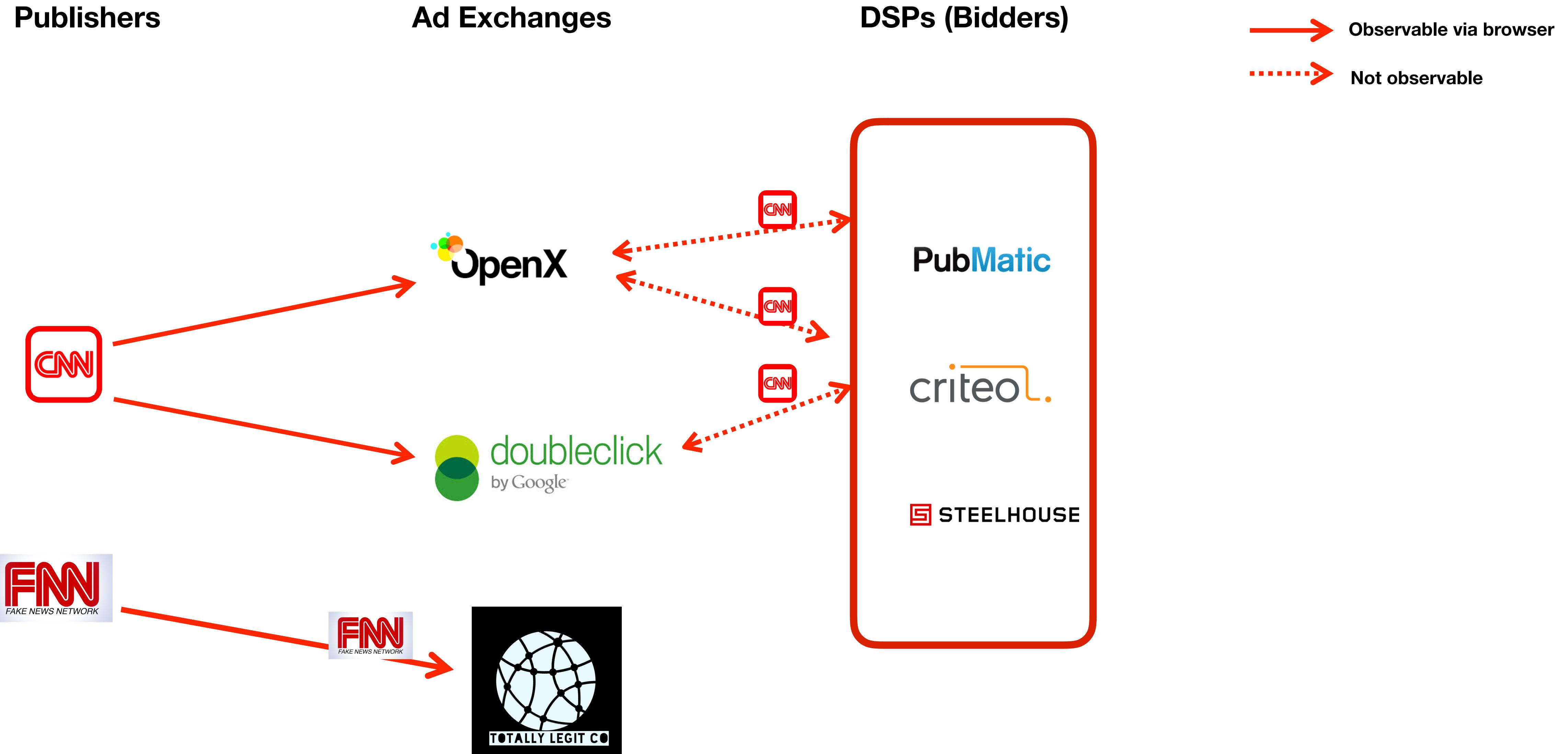
Potential for Fraud in RTB



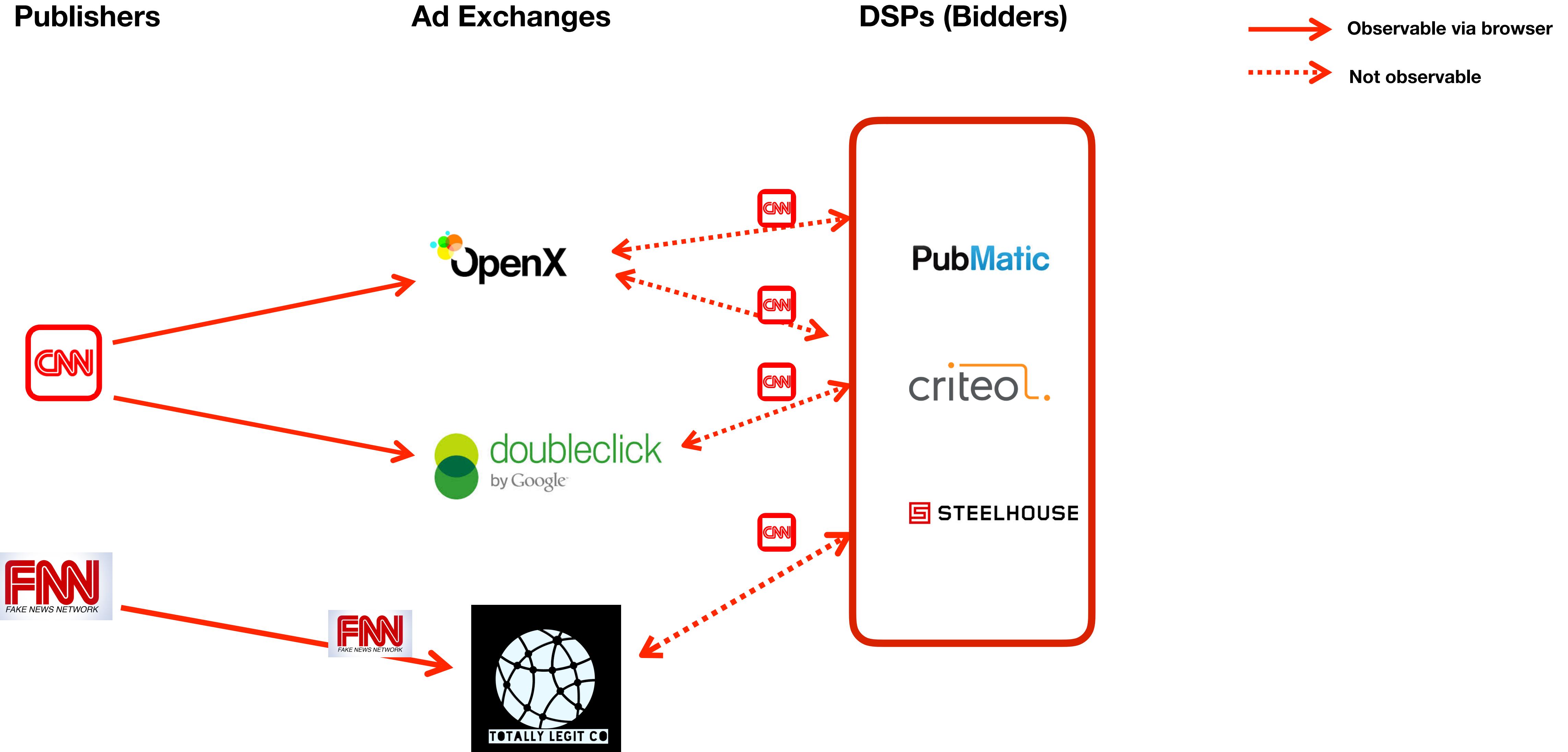
Potential for Fraud in RTB



Potential for Fraud in RTB



Potential for Fraud in RTB



Ad Fraud = Loss in \$\$\$

Ad Fraud = Loss in \$\$\$

AdAge

FCB, BBDO moves Dentsu forecasts NYT ad sales McDonald's cups

Digital

AD FRAUD WILL COST \$7.2 BILLION IN 2016, ANA SAYS, UP NEARLY \$1 BILLION

Study Recommends Better Understanding of Programmatic Supply Chain

By George P. Scott

The \$8.2 Billion Adtech Fraud Problem That Everyone Is Ignoring

Samuel Scott 4 years ago Comment



Talk to an enthusiast

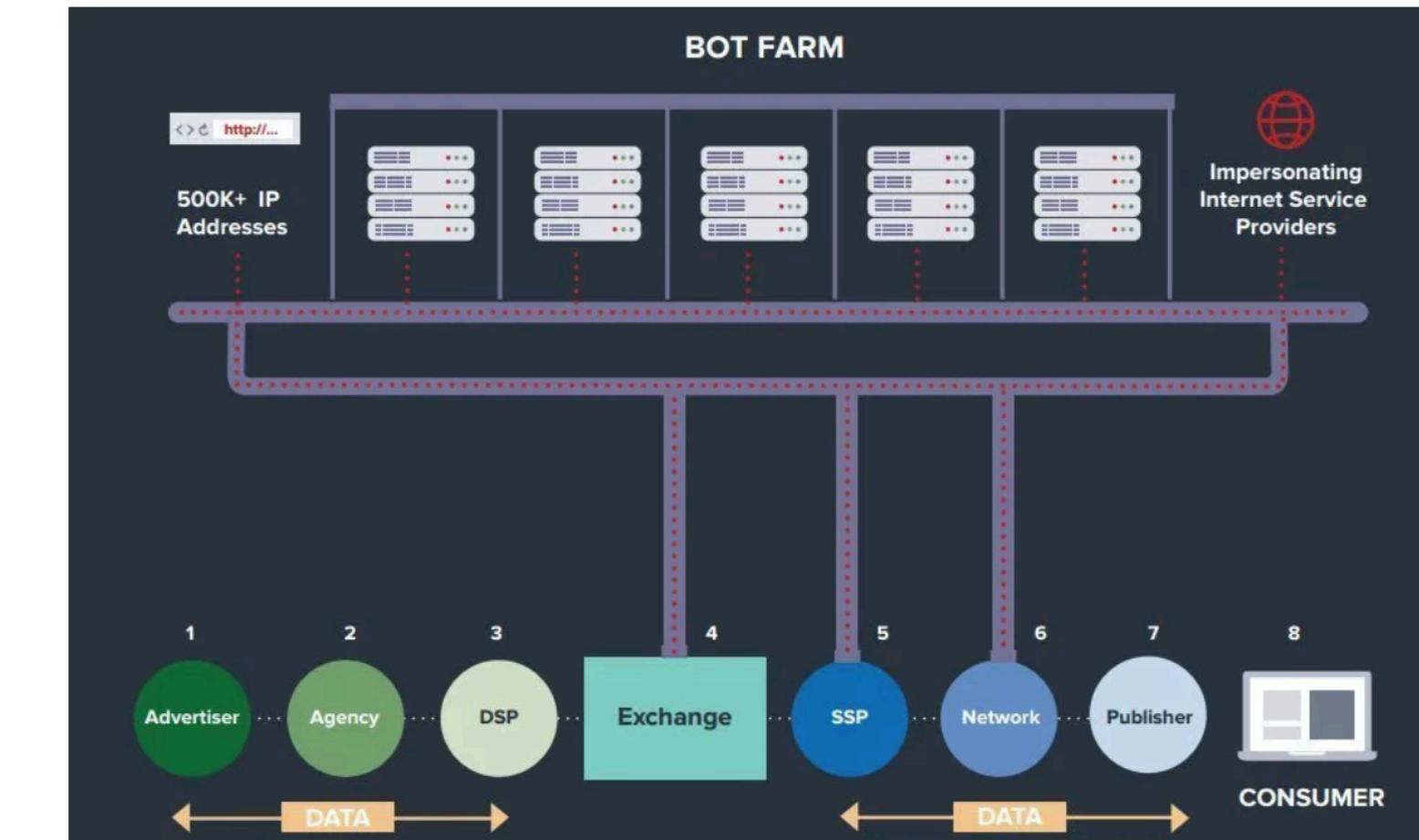
Startups
Apps
Gadgets
Videos
Audio
Extra Crunch NEW
Newsletters

Events
Advertise
Crunchbase
More

Search

Facebook privacy
Google
Transportation
Enterprise

Methbot: a \$3M-\$5M/day video ad-tech fraud



White Ops, a security firm, has published a detailed report on a crime-ring they call "Methbot" that generated \$3M-\$5M by creating 6,000 fake websites to embed videos in, then generating convincing bots that appeared to

THE PROGRAMMATIC MARKETER

Domain spoofing remains a huge threat to programmatic

FEBRUARY 28, 2017 by Yuyu Chen



Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

<https://cnn.com/ads.txt>

```
.....  
google.com, pub-7439281311086140, DIRECT  
rubiconproject.com, 11078, DIRECT  
c.amazon-adsystem.com, 3159, DIRECT  
openx.com, 183753, RESELLER  
.....
```

Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

<https://cnn.com/ads.txt>

```
.....  
google.com, pub-7439281311086140, DIRECT  
rubiconproject.com, 11078, DIRECT  
c.amazon-adsystem.com, 3159, DIRECT  
openx.com, 183753, RESELLER  
.....
```

File Format:

<FIELD #1>, <FIELD #2>, <FIELD #3>

#1: Exchange domain *

#2: Publisher's account ID

#3: Type of account/relationship

Programmatic Buying & Selling via ads.txt Standard

Authorized Digital Sellers (ads.txt)

- Introduced by Interactive Advertising Bureau (IAB) in May 2017
- Enables publishers to list authorized exchanges
- At the root of the top-level domain

<https://cnn.com/ads.txt>

```
.....  
google.com, pub-7439281311086140, DIRECT  
rubiconproject.com, 11078, DIRECT  
c.amazon-adsystem.com, 3159, DIRECT  
openx.com, 183753, RESELLER  
.....
```

File Format:

<FIELD #1>, <FIELD #2>, <FIELD #3>

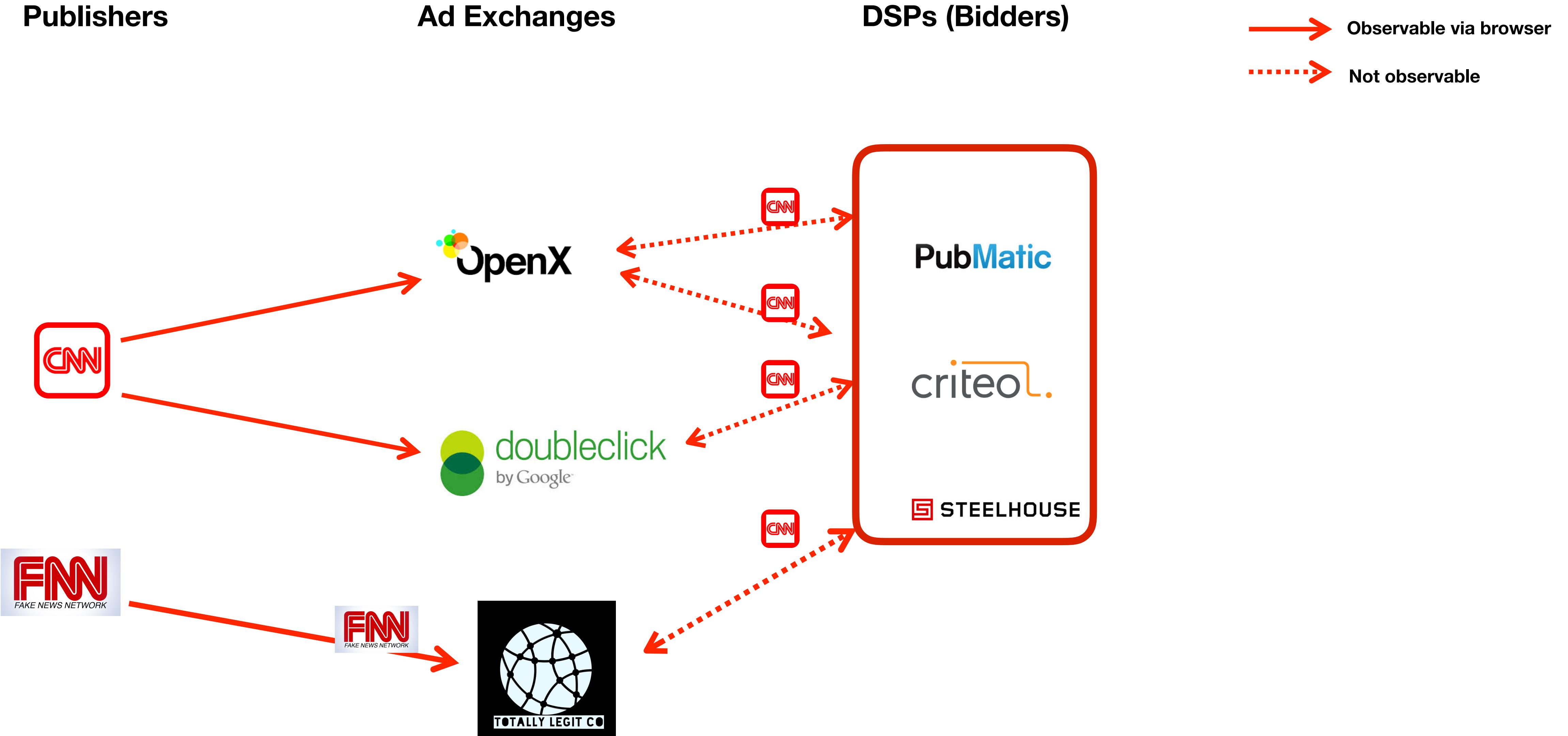
#1: Exchange domain *

#2: Publisher's account ID

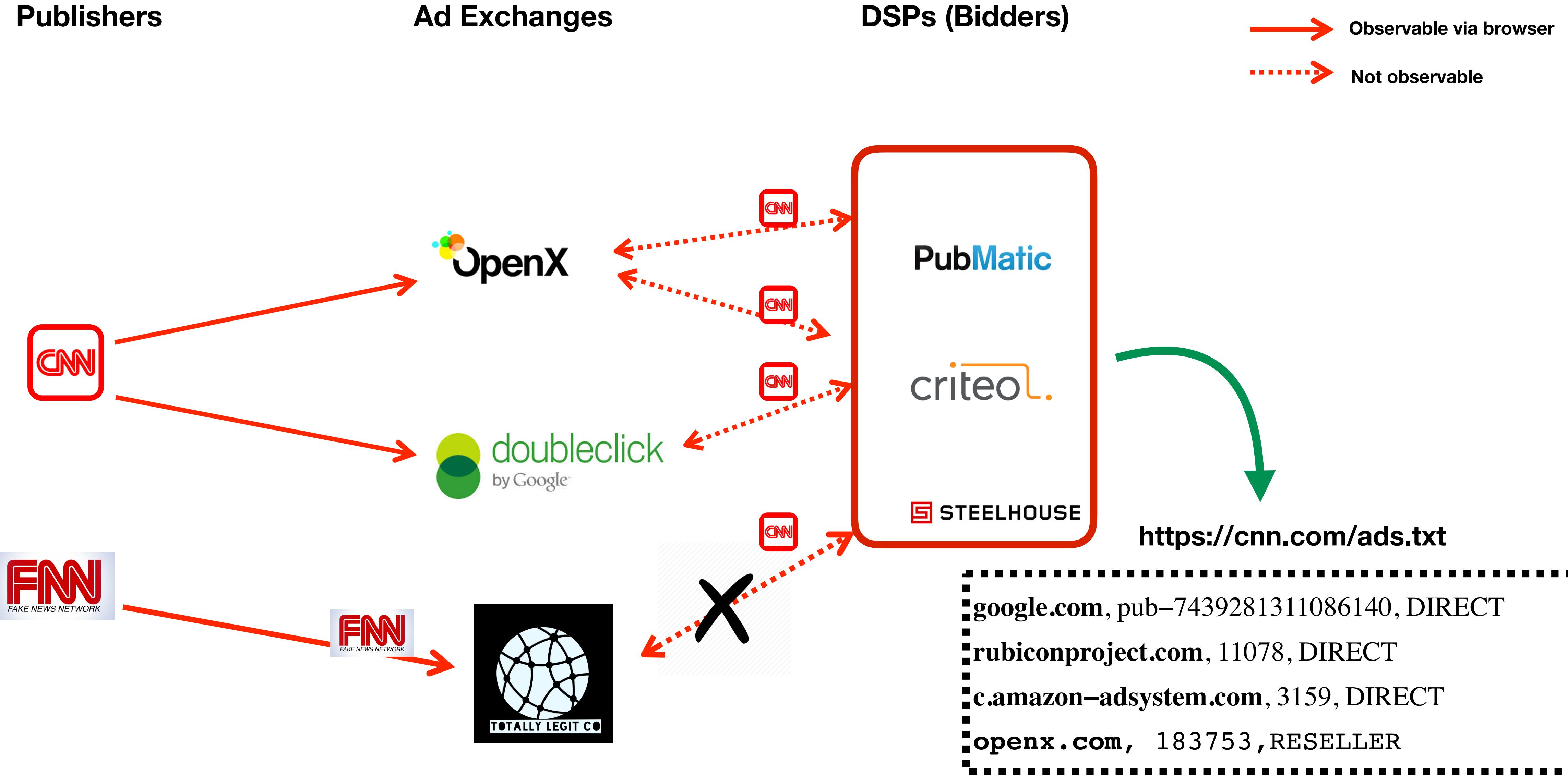
#3: Type of account/relationship

* Not always an exchange

Potential for Fraud in RTB



Potential for Fraud in RTB



Opportunity to Identify Ad Exchanges

<https://cnn.com/ads.txt>

```
.....  
google.com, pub-7439281311086140, DIRECT  
rubiconproject.com, 11078, DIRECT  
c.amazon-adsystem.com, 3159, DIRECT  
openx.com, 183753, RESELLER
```

<https://washingtonpost.com/ads.txt>

```
.....  
indexexchange.com, 197545, DIRECT  
adtech.com, 232547, DIRECT  
openx.com, 203799, RESELLER
```

We can extract FIELD#1 from publishers to
isolate a list of Ad exchange

Methodology for Collecting ads.txt Files

Crawled Alexa Top 100K websites

- For each domain, fetched /ads.txt file
- Used Python's *requests* module
- 26 snapshots between January 2018 – April 2019

Methodology for Collecting ads.txt Files

Crawled Alexa Top 100K websites

- For each domain, fetched /ads.txt file
- Used Python's *requests* module
- 26 snapshots between January 2018 – April 2019

<https://cnn.com/ads.txt>

```
-----  
google.com, pub-7439281311086140, DIRECT  
rubiconproject.com, 11078, DIRECT  
c.amazon-adsystem.com, 3159, DIRECT  
openx.com, 183753, RESELLER RESELLER  
-----
```

<https://washingtonpost.com/ads.txt>

```
-----  
indexexchange.com, 197545, DIRECT  
adtech.com, 232547, DIRECT  
openx.com, 203799, RESELLER  
-----
```

Methodology for Collecting ads.txt Files

Crawled Alexa Top 100K websites

- For each domain, fetched /ads.txt file
- Used Python's *requests* module
- 26 snapshots between January 2018 – April 2019

<https://cnn.com/ads.txt>

```
-----  
google.com, pub-7439281311086140, DIRECT  
rubiconproject.com, 11078, DIRECT  
c.amazon-adsystem.com, 3159, DIRECT  
openx.com, 183753, RESELLER RESELLER  
-----
```

<https://washingtonpost.com/ads.txt>

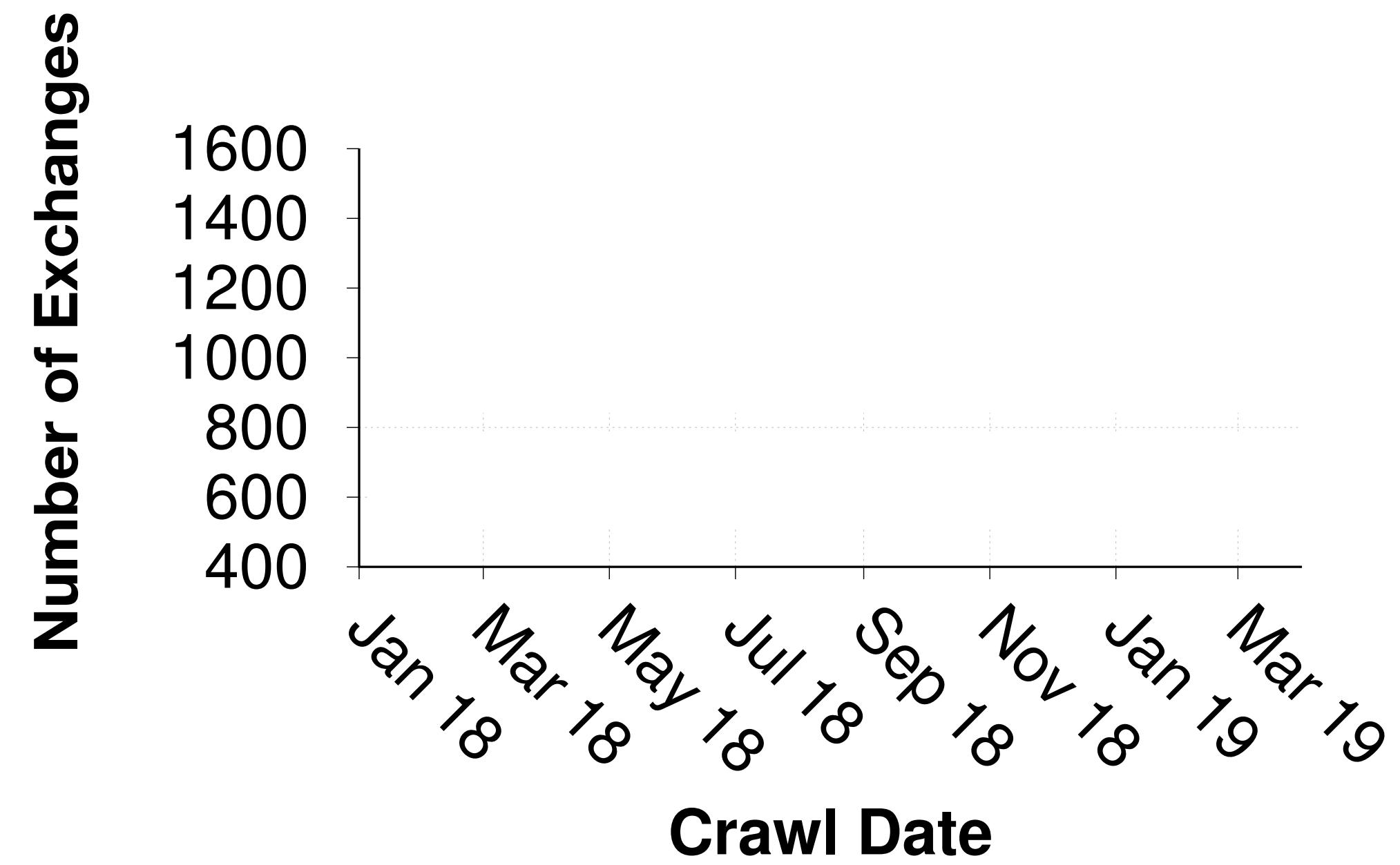
```
-----  
indexexchange.com, 197545, DIRECT  
adtech.com, 232547, DIRECT  
openx.com, 203799, RESELLER  
-----
```

Extracted first comma-separated value (exchange domain)

```
google.com  
rubiconproject.com  
c.amazon-adsystem.com  
openx.com
```

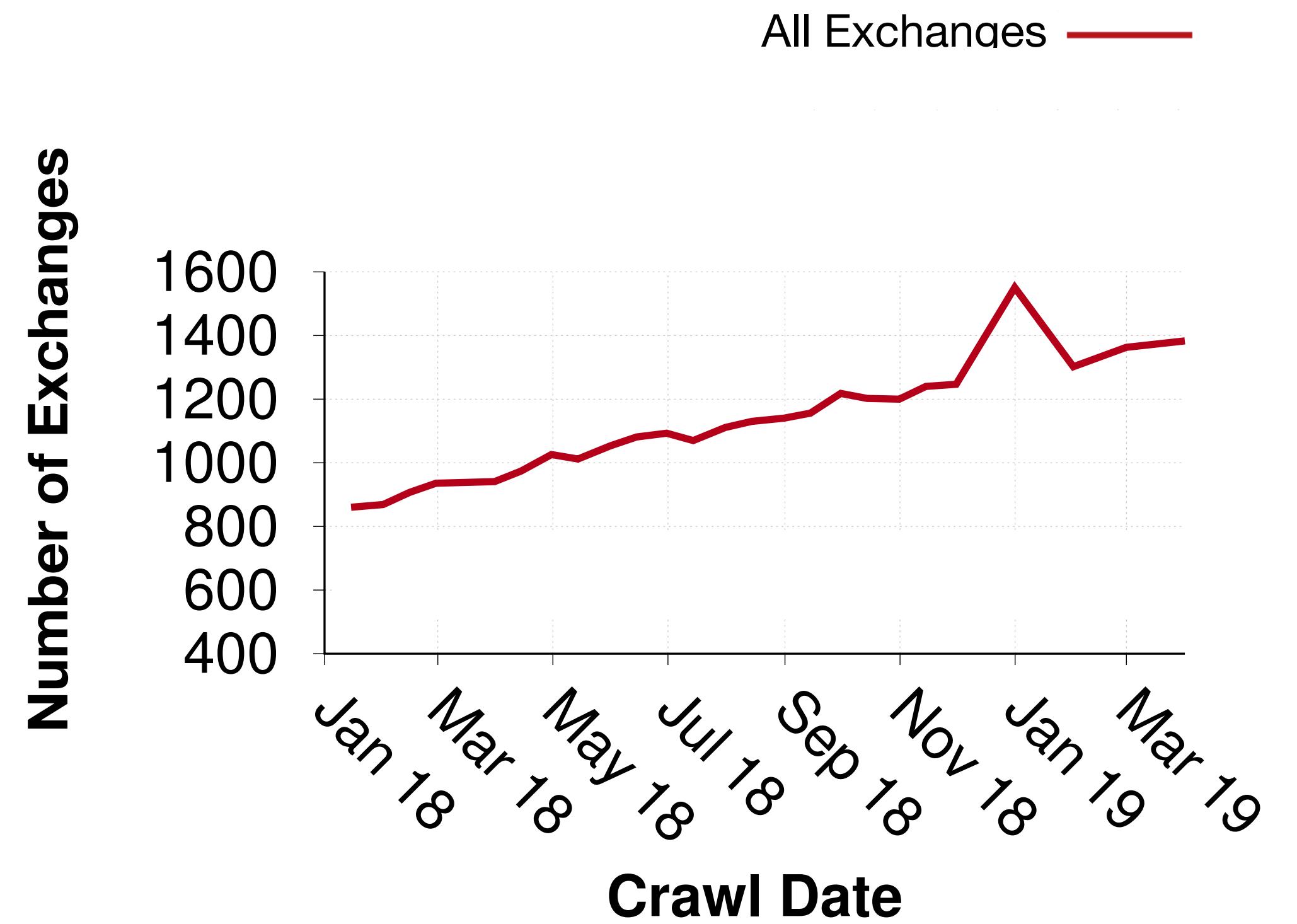
```
indexexchange.com  
adtech.com  
openx.com
```

Authorized Exchanges Discovered



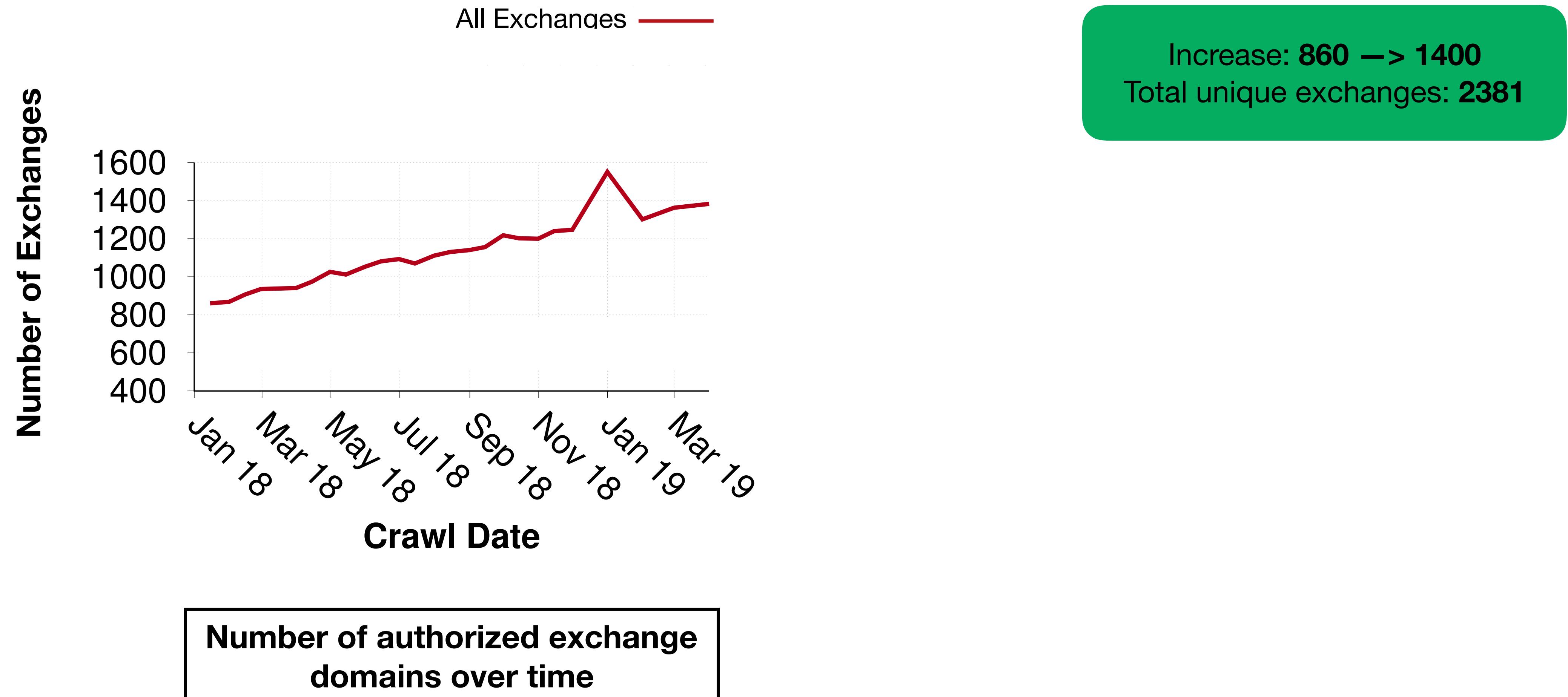
**Number of authorized exchange
domains over time**

Authorized Exchanges Discovered

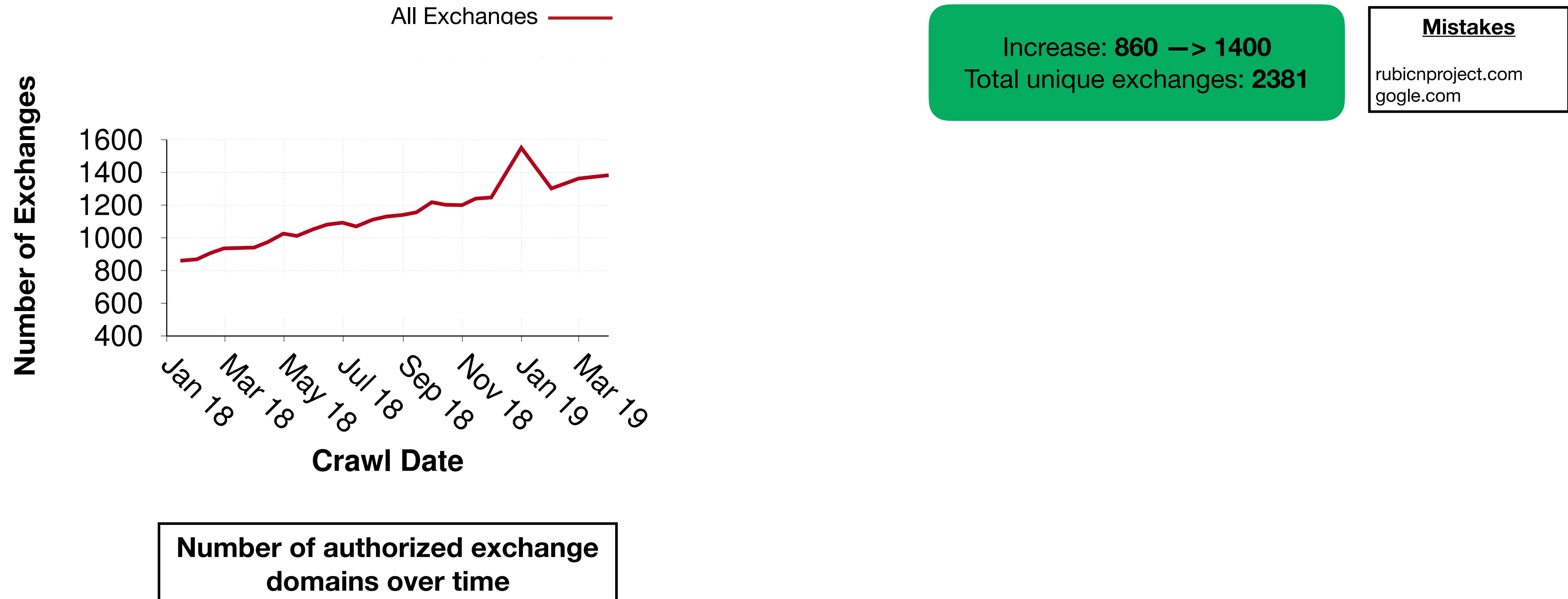


**Number of authorized exchange
domains over time**

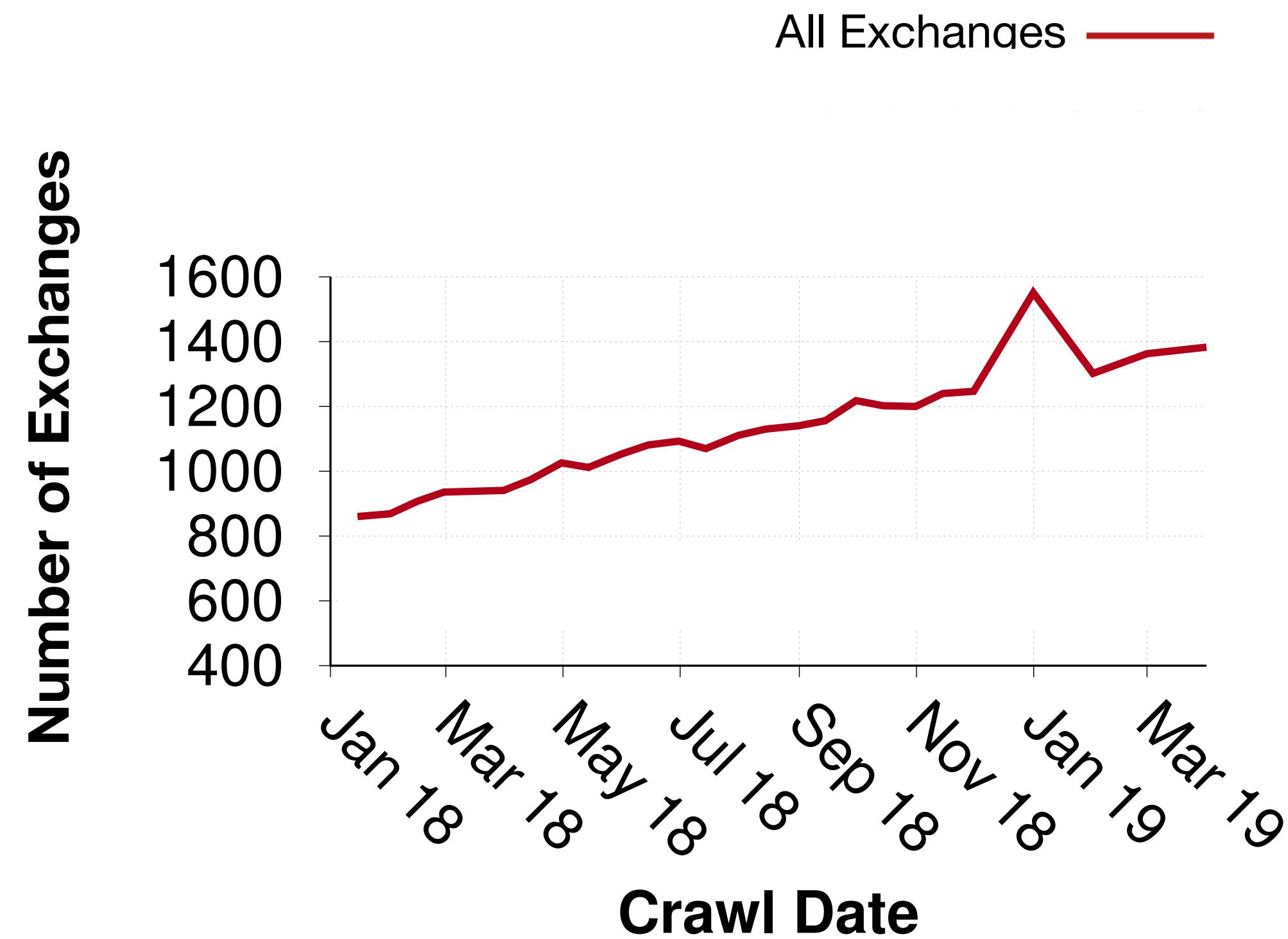
Authorized Exchanges Discovered



Authorized Exchanges Discovered



Authorized Exchanges Discovered



Number of authorized exchange domains over time

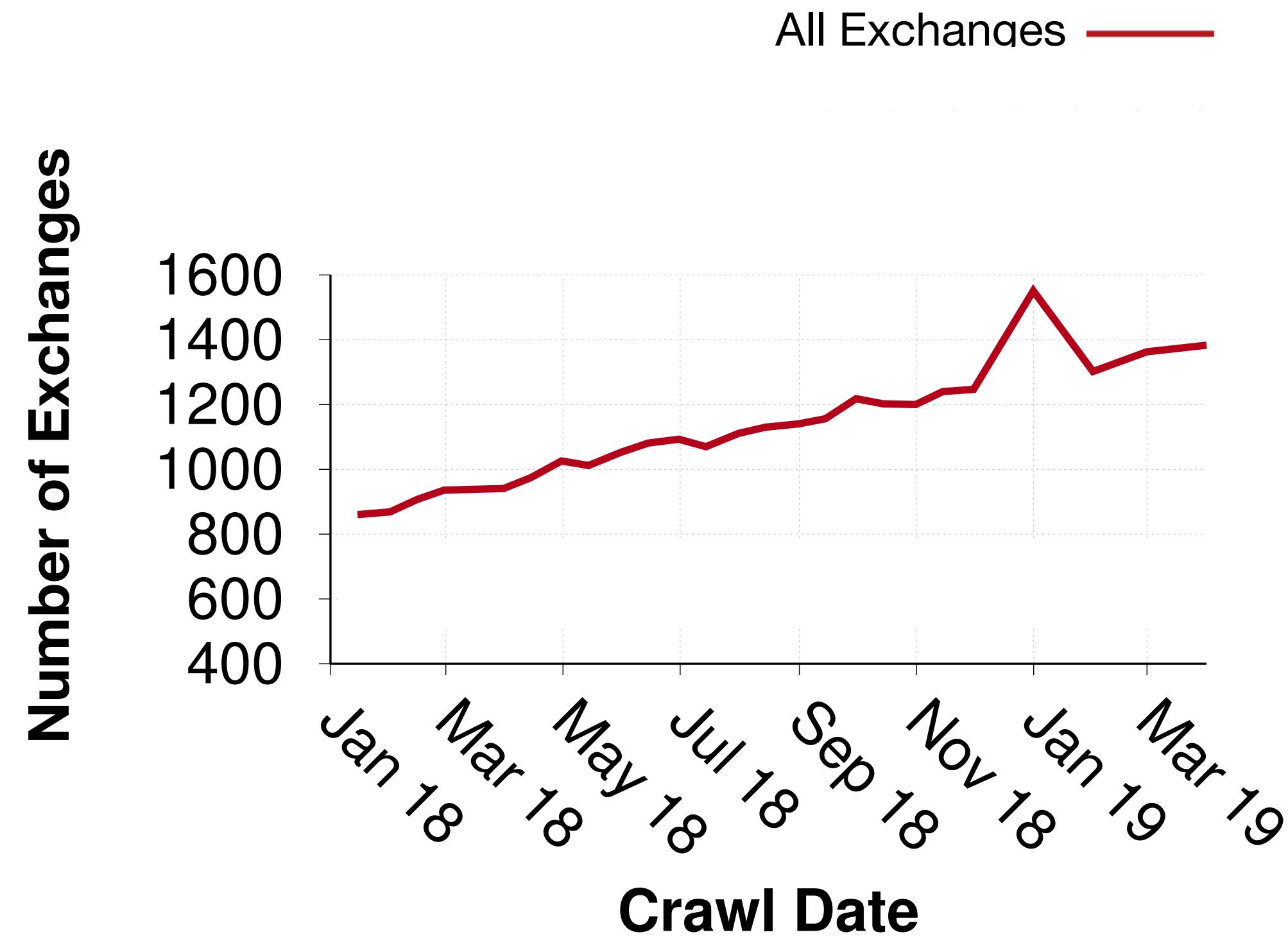
Increase: 860 → 1400
Total unique exchanges: 2381

Fetch WHOIS record

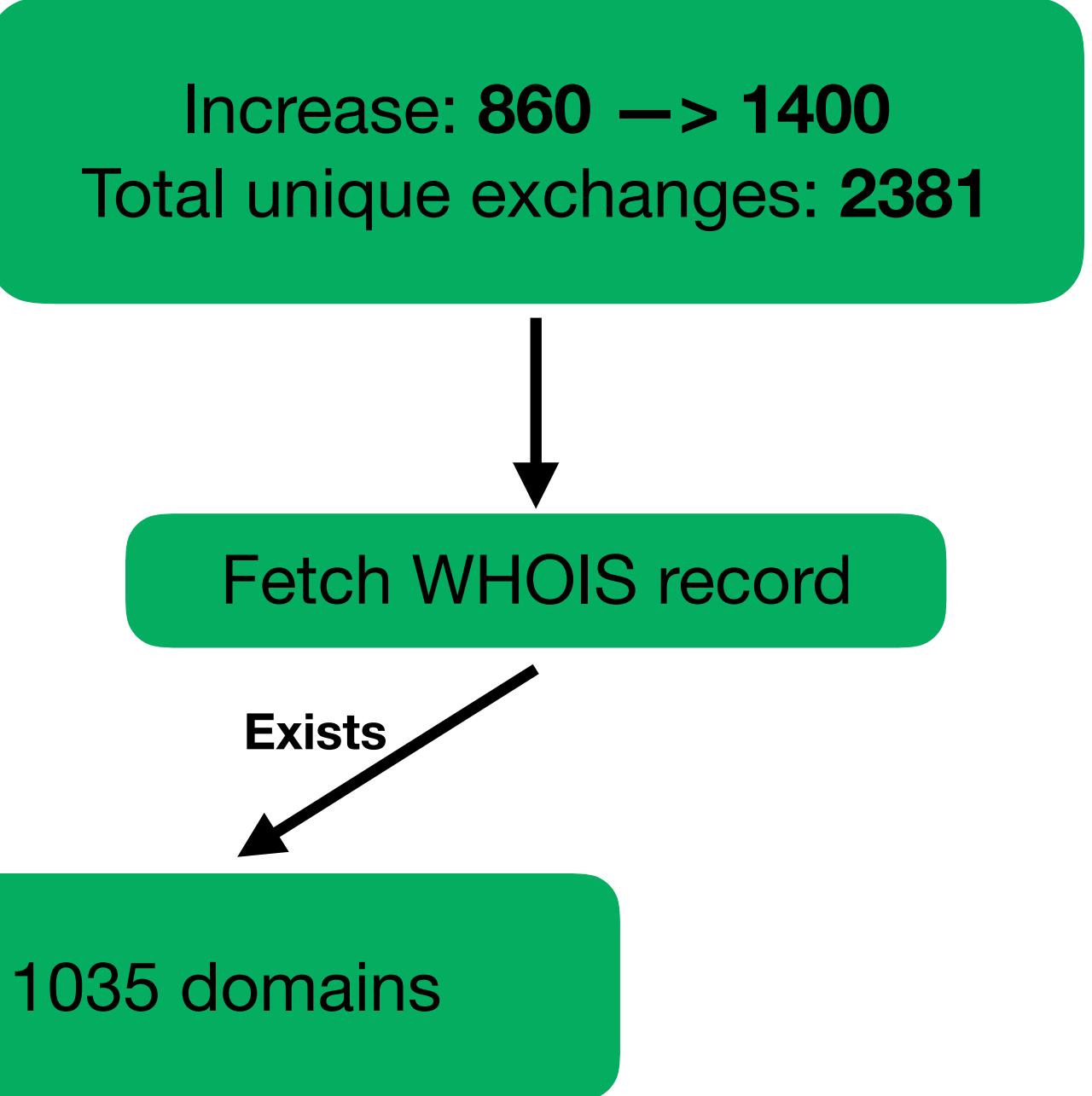
Mistakes

rubicnproject.com
gogle.com

Authorized Exchanges Discovered



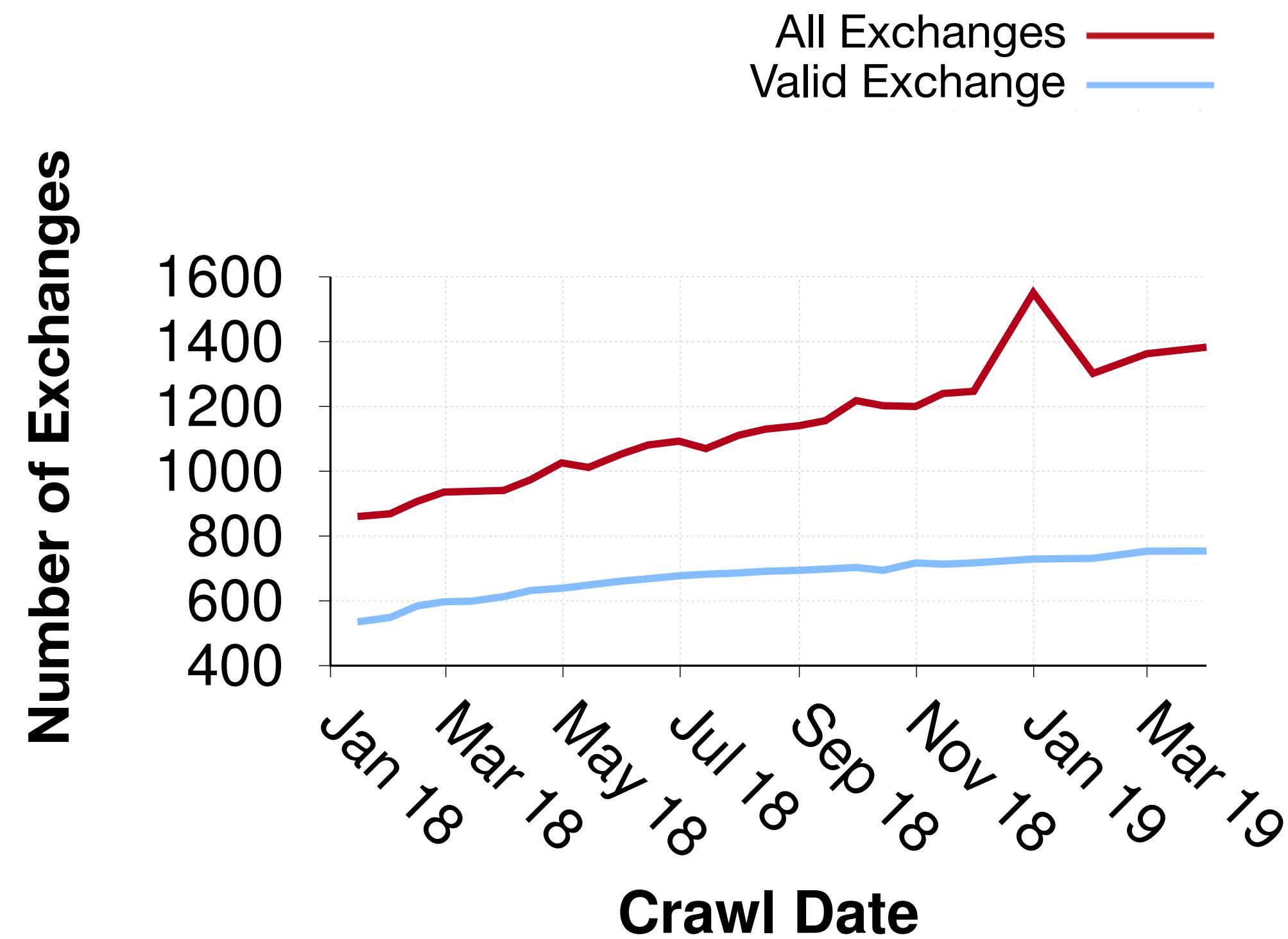
Number of authorized exchange domains over time



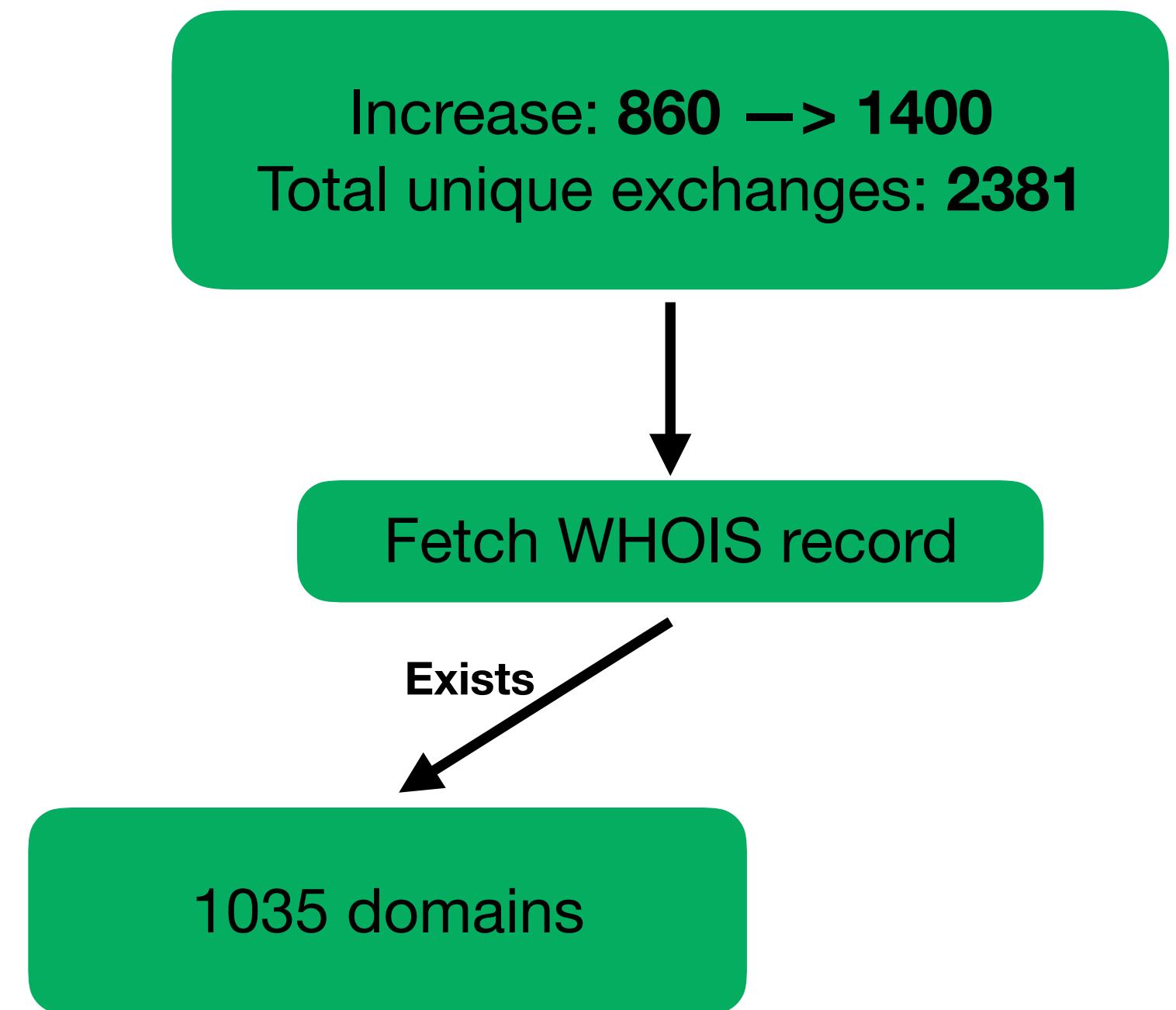
Mistakes

rubicnproject.com
gogle.com

Authorized Exchanges Discovered



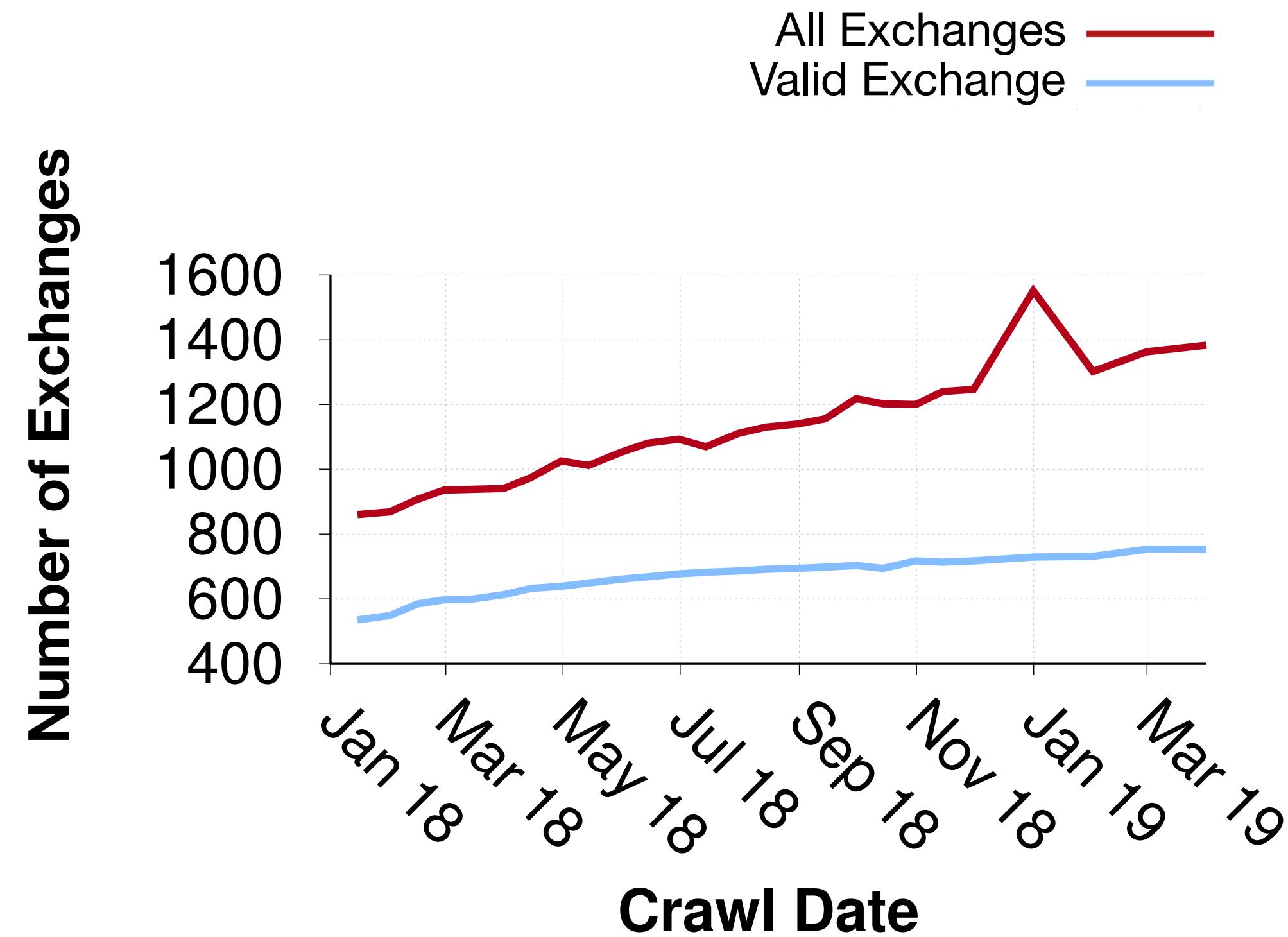
Number of authorized exchange domains over time



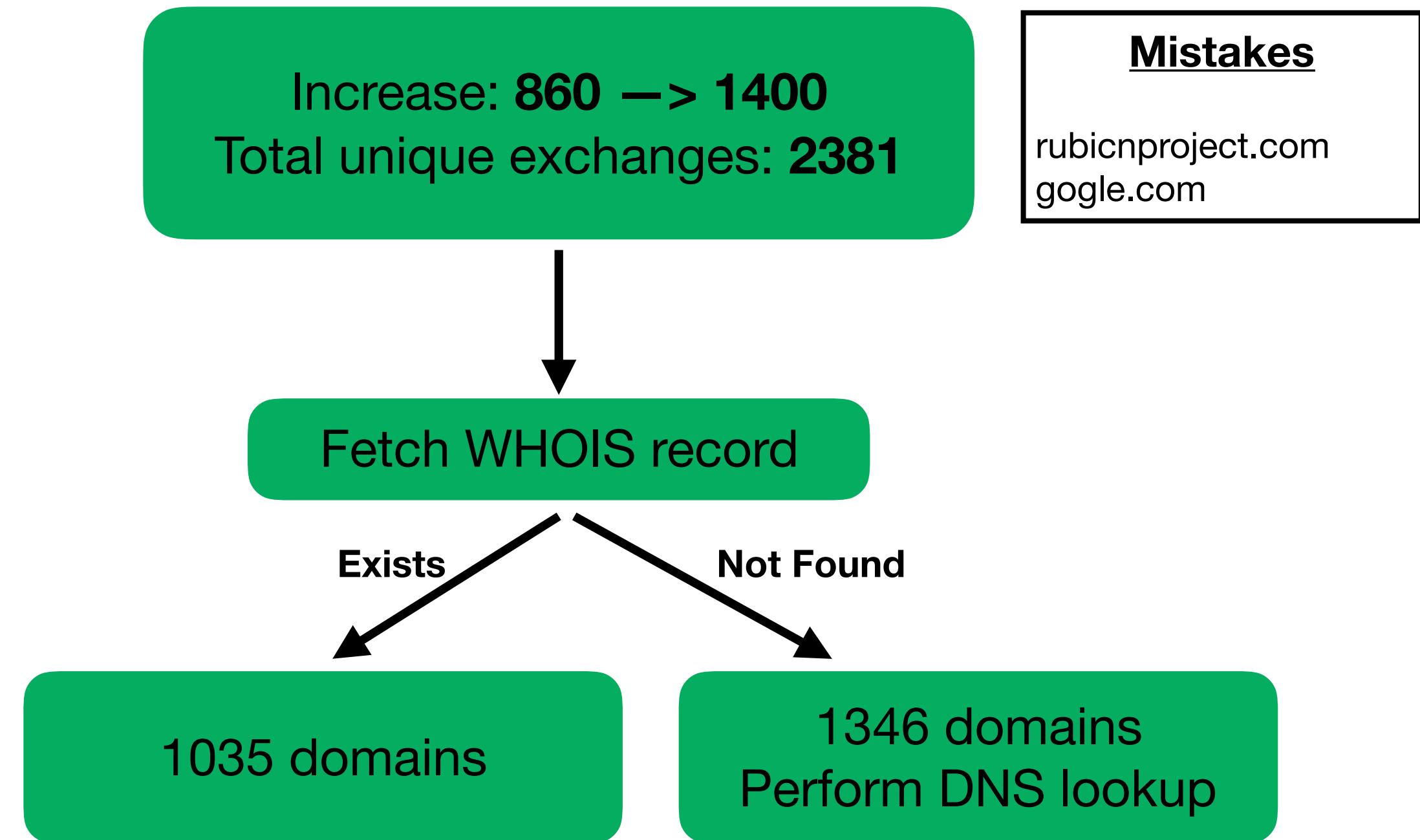
Mistakes

rubicnproject.com
gogle.com

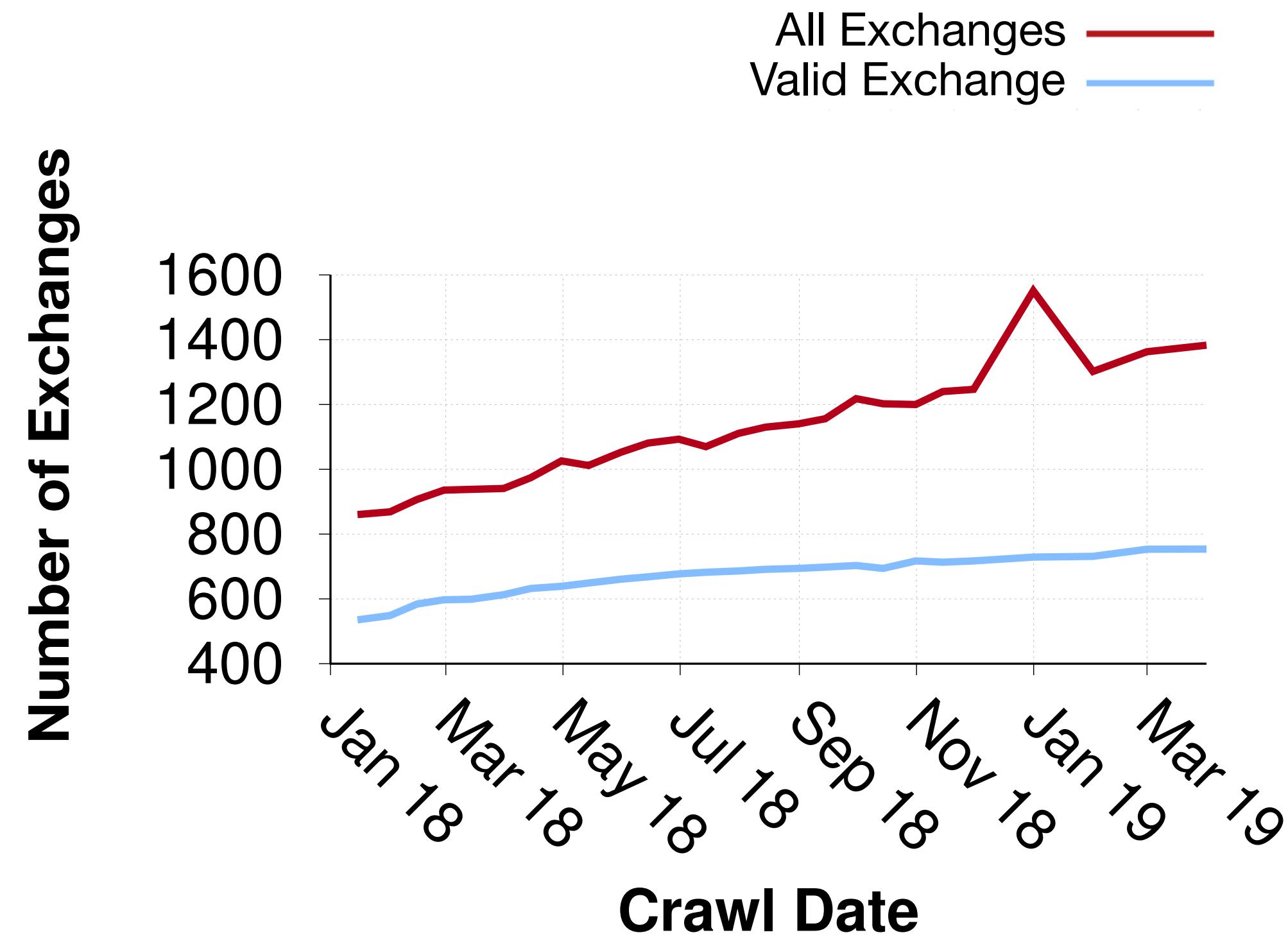
Authorized Exchanges Discovered



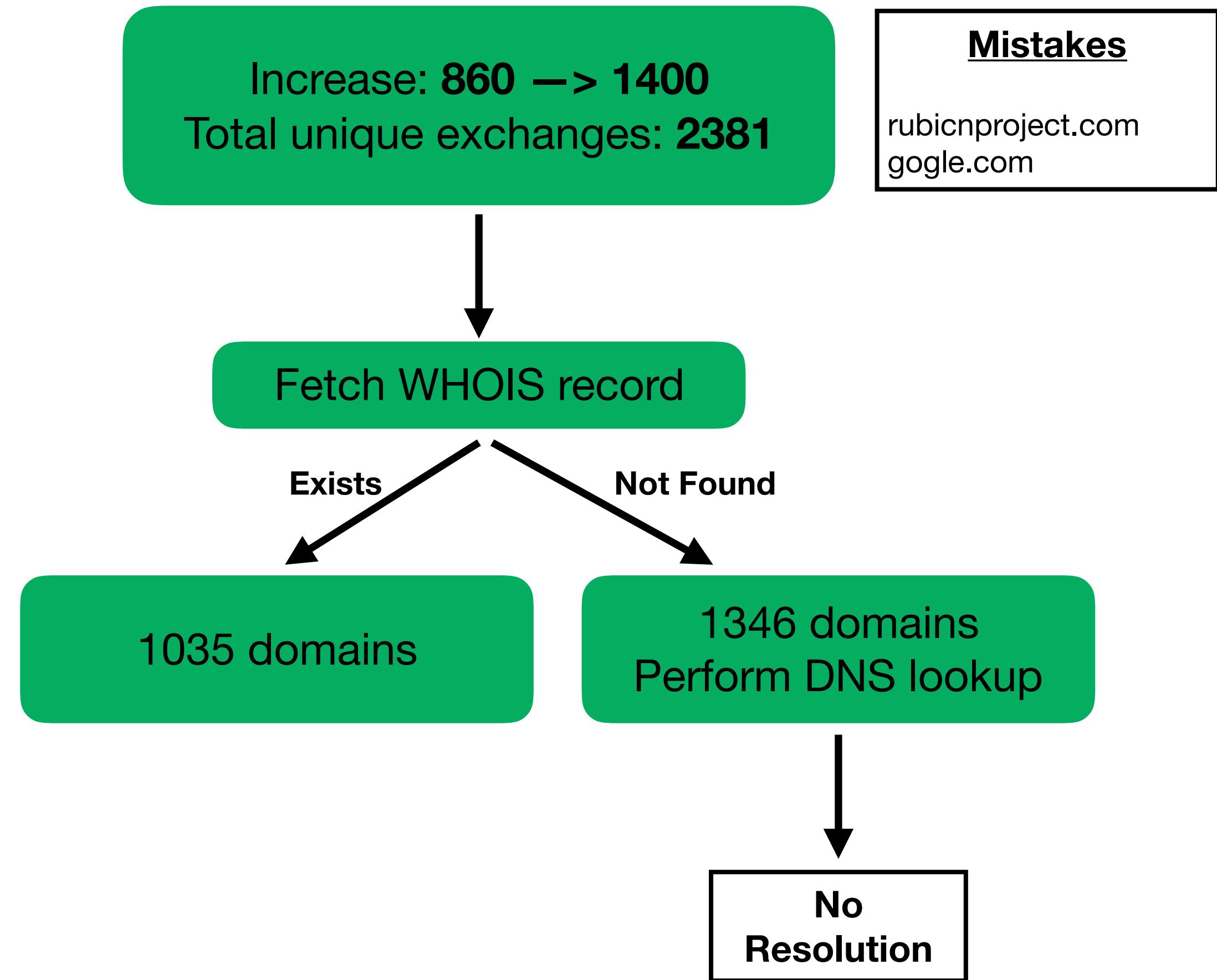
Number of authorized exchange domains over time



Authorized Exchanges Discovered



Number of authorized exchange domains over time



More Results in Paper

A Longitudinal Analysis of the ads.txt Standard

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson. [IMC 2019]

- Detailed adoption trends by publishers and exchanges
- Compliance of the standard by DSPs and exchanges

Is This Data Representative For Our Purposes?

Is This Data Representative For Our Purposes?

Q: Does ads.txt have good coverage?

- 60% of the publishers from Alexa top-100K who run RTB ads.
- Major A&A domains support ads.txt



Is This Data Representative For Our Purposes?

Q: Does ads.txt have good coverage?

- 60% of the publishers from Alexa top-100K who run RTB ads.
- Major A&A domains support ads.txt



Q: Is the data collected clean?

- Some manual intervention is necessary
- For example: publishers added **google.com**, but the exchange is **doubleclick**

Thesis Outline

Detecting Information Sharing

Isolating List of Ad Exchanges

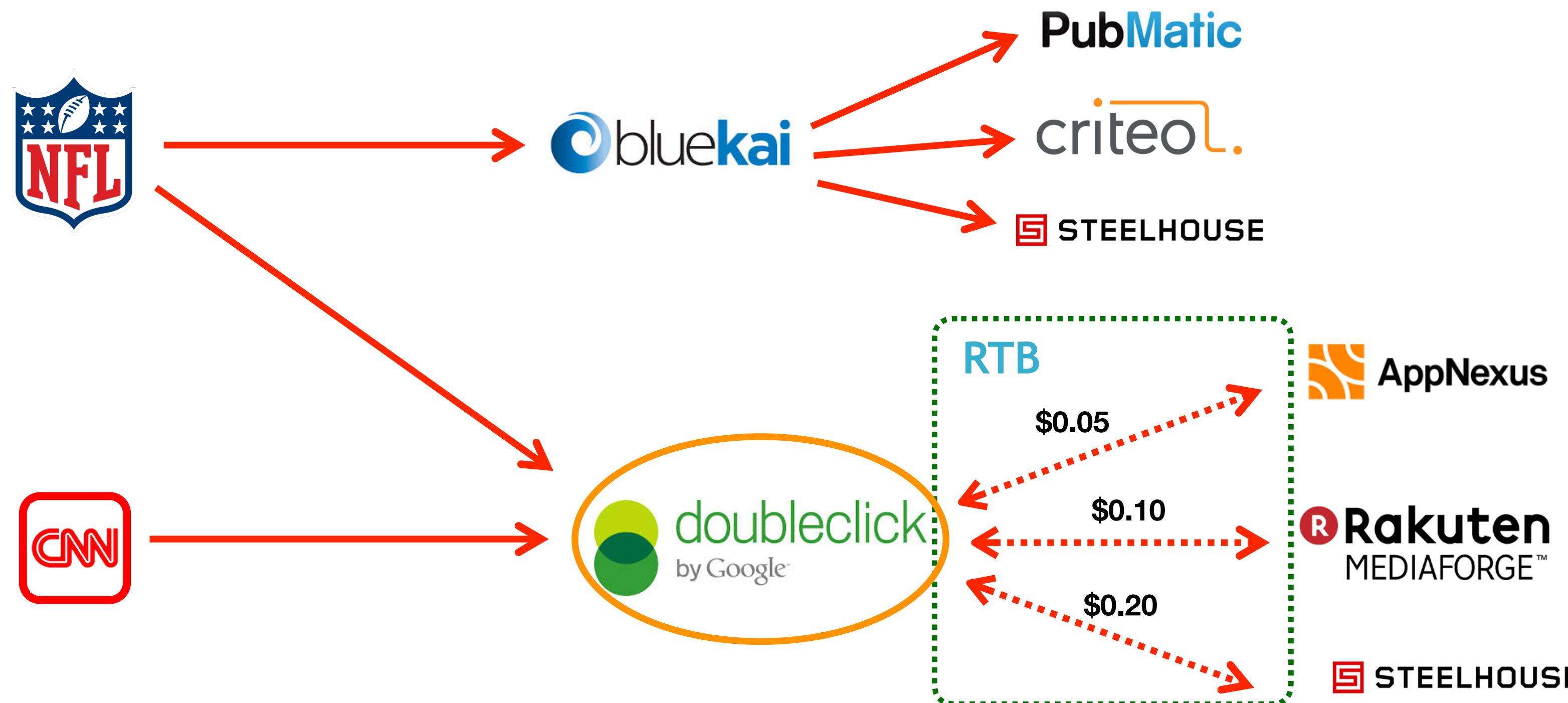
Modeling User's Privacy Digital Footprint

What Have We Detected So Far?

Publishers

Advertising & Analytics (A&A) companies

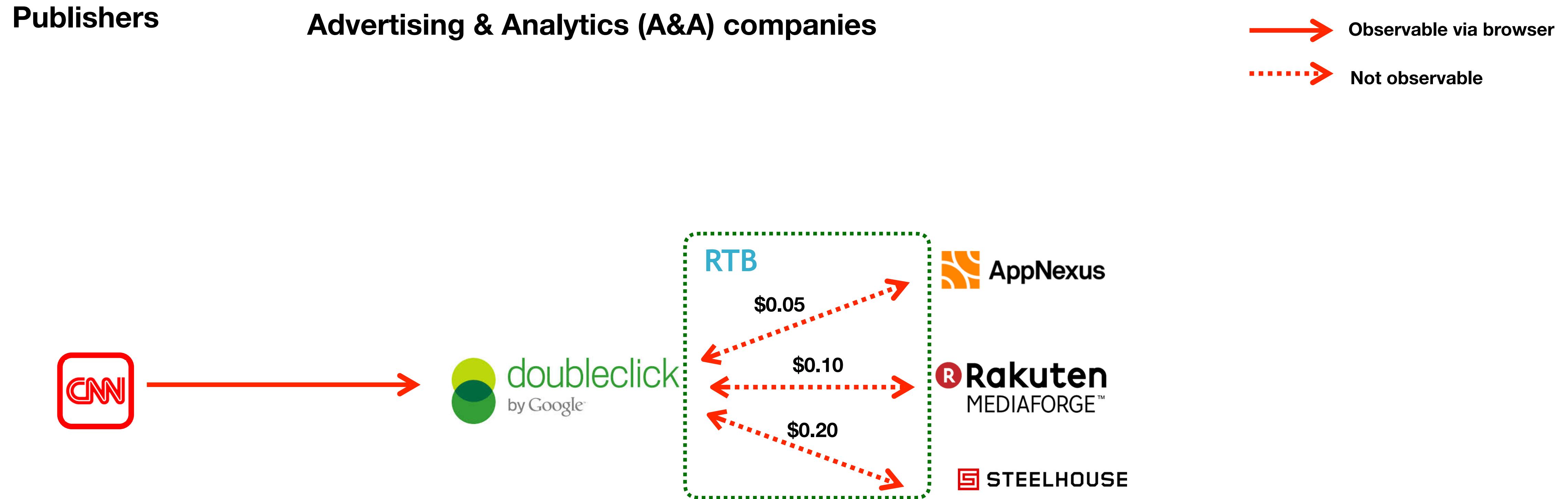
→ Observable via browser
→ Not observable



Need to **accurately** capture information-sharing relationships.

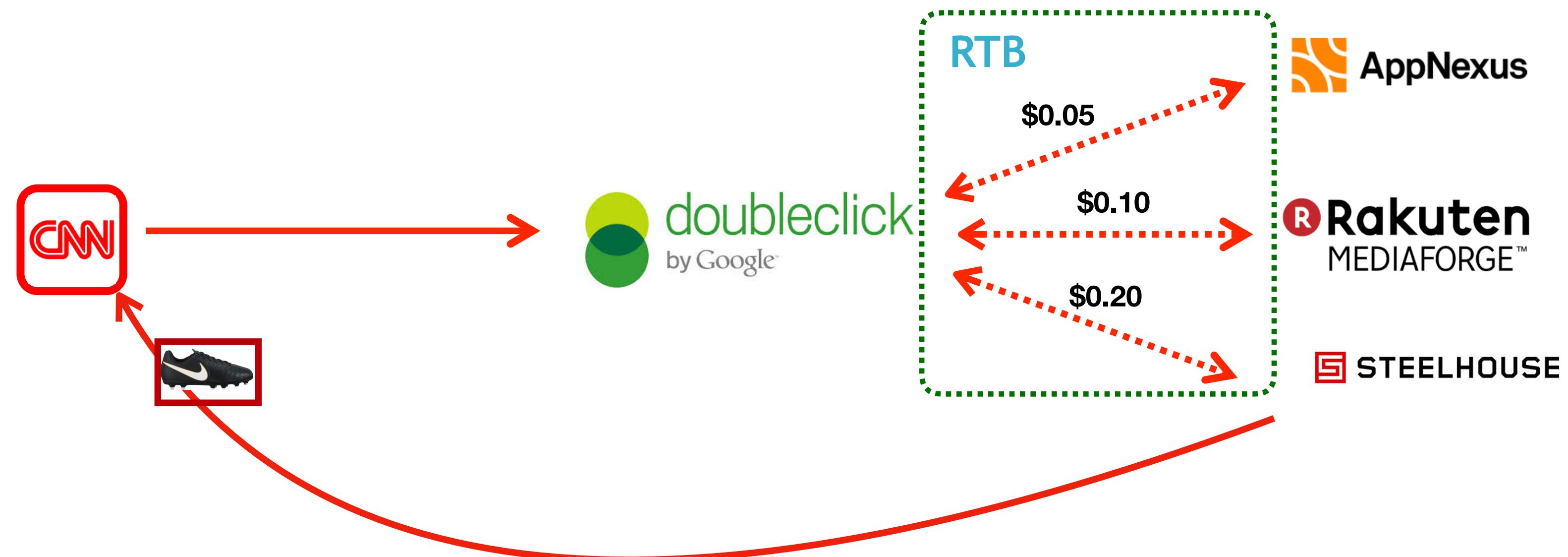
Identify A&A companies that hold RTB auctions

Enumerating DSP Partners of Ad Exchanges

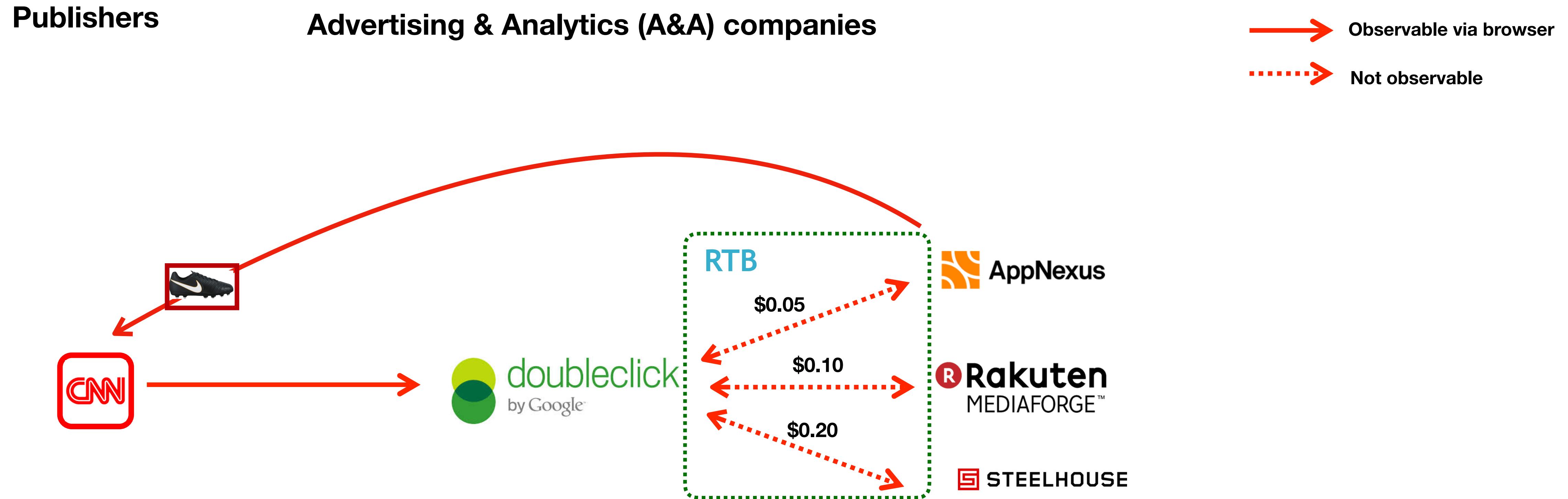


Enumerating DSP Partners of Ad Exchanges

Publishers → Observable via browser
Advertising & Analytics (A&A) companies → Not observable



Enumerating DSP Partners of Ad Exchanges



Goal & Research Questions

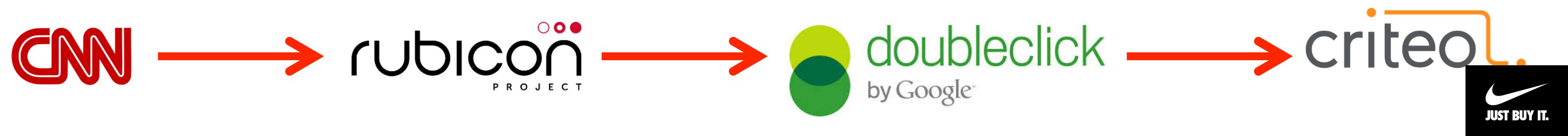
Model the Diffusion of Impressions in the Advertising Ecosystem

Taking RTB into account

Key Questions:

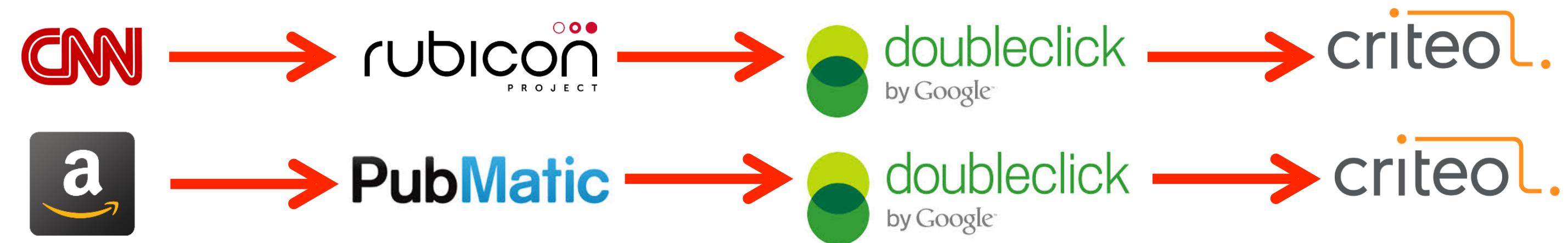
1. Which A&A domains see user impressions?
2. What fraction of user impressions are viewed by A&A domains?
3. How much ad and tracker blocking extensions help?

Recap - Inclusion Chains

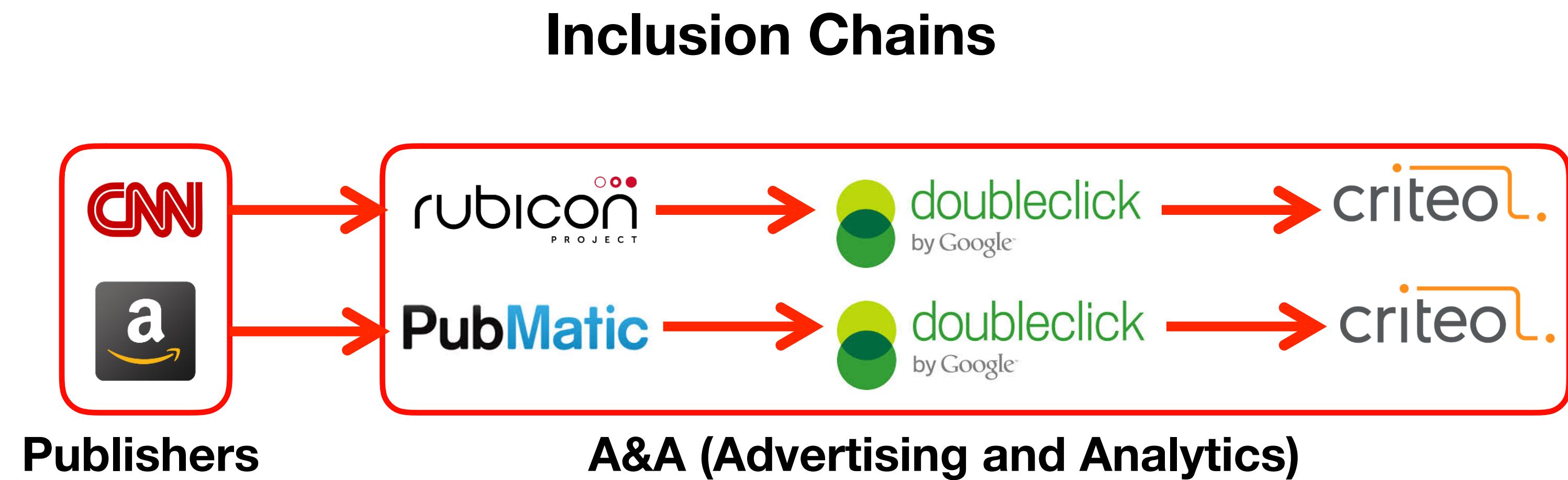


From Chains to Graph

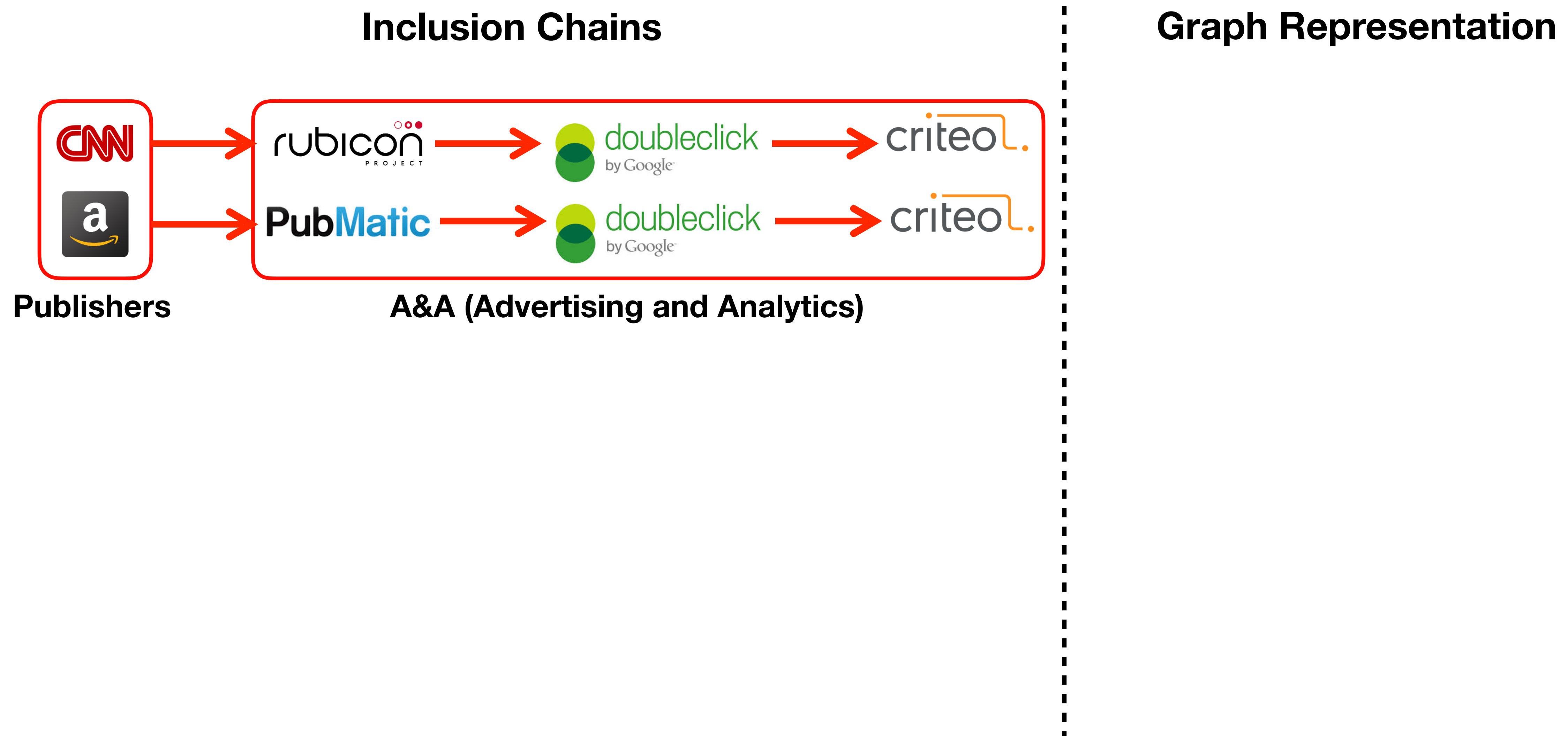
Inclusion Chains



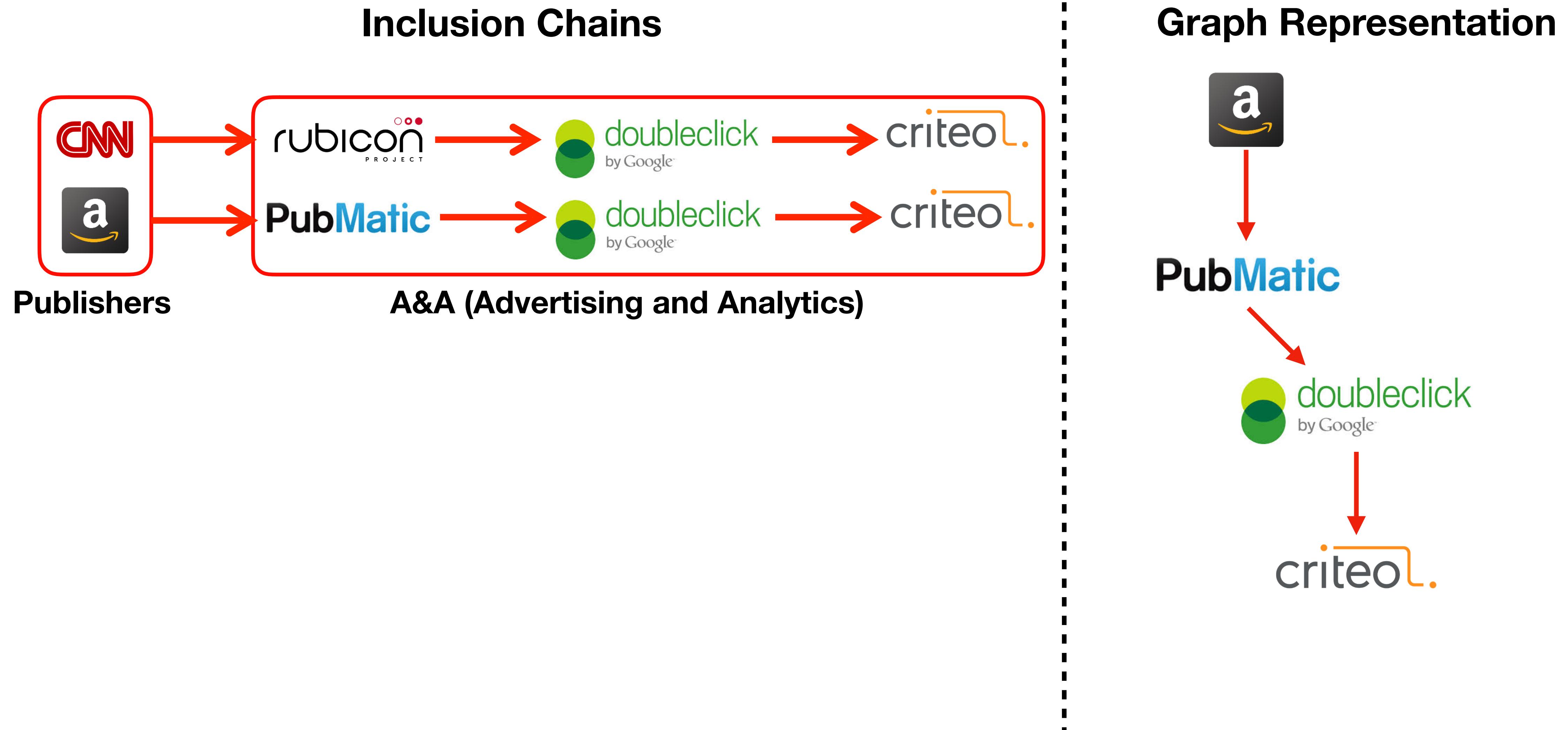
From Chains to Graph



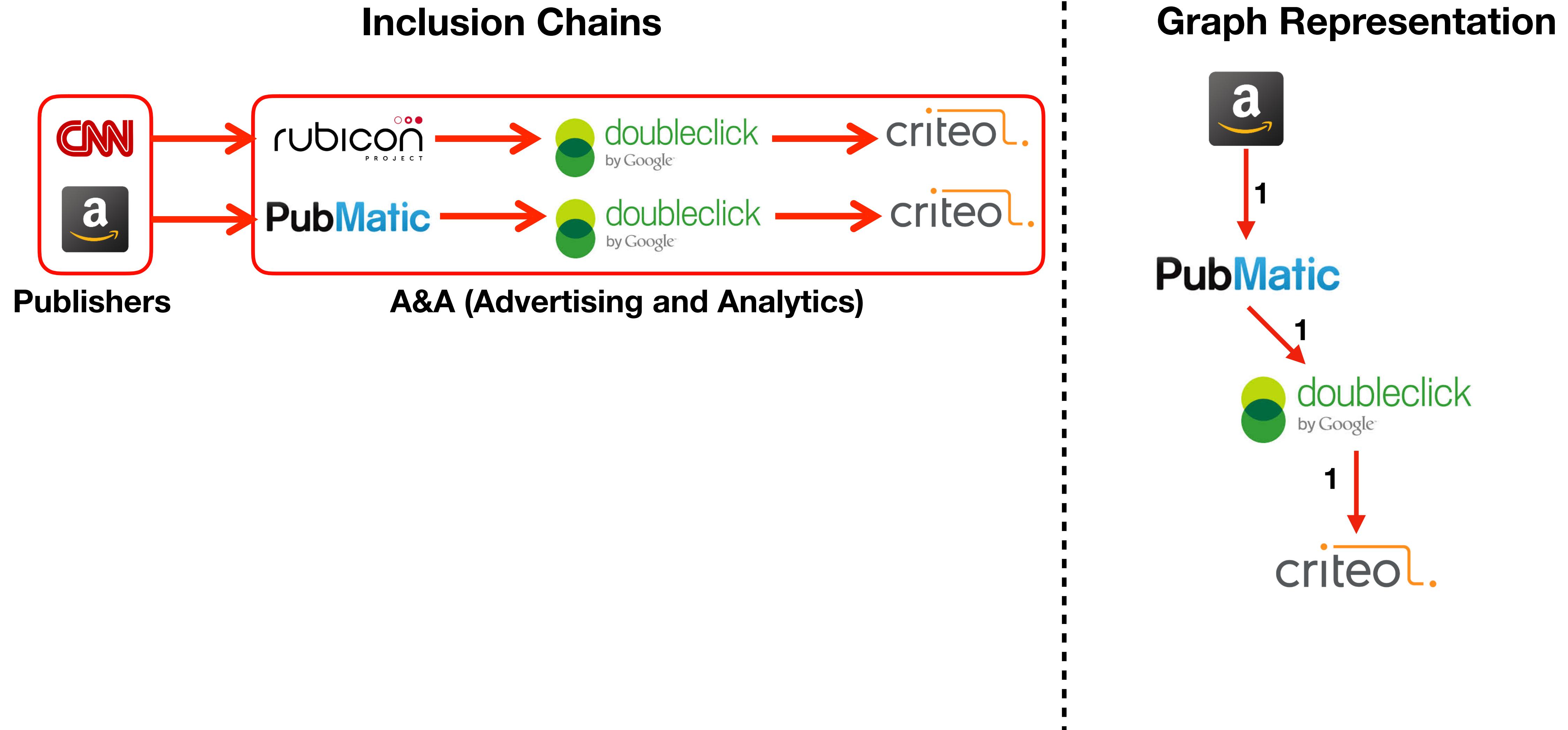
From Chains to Graph



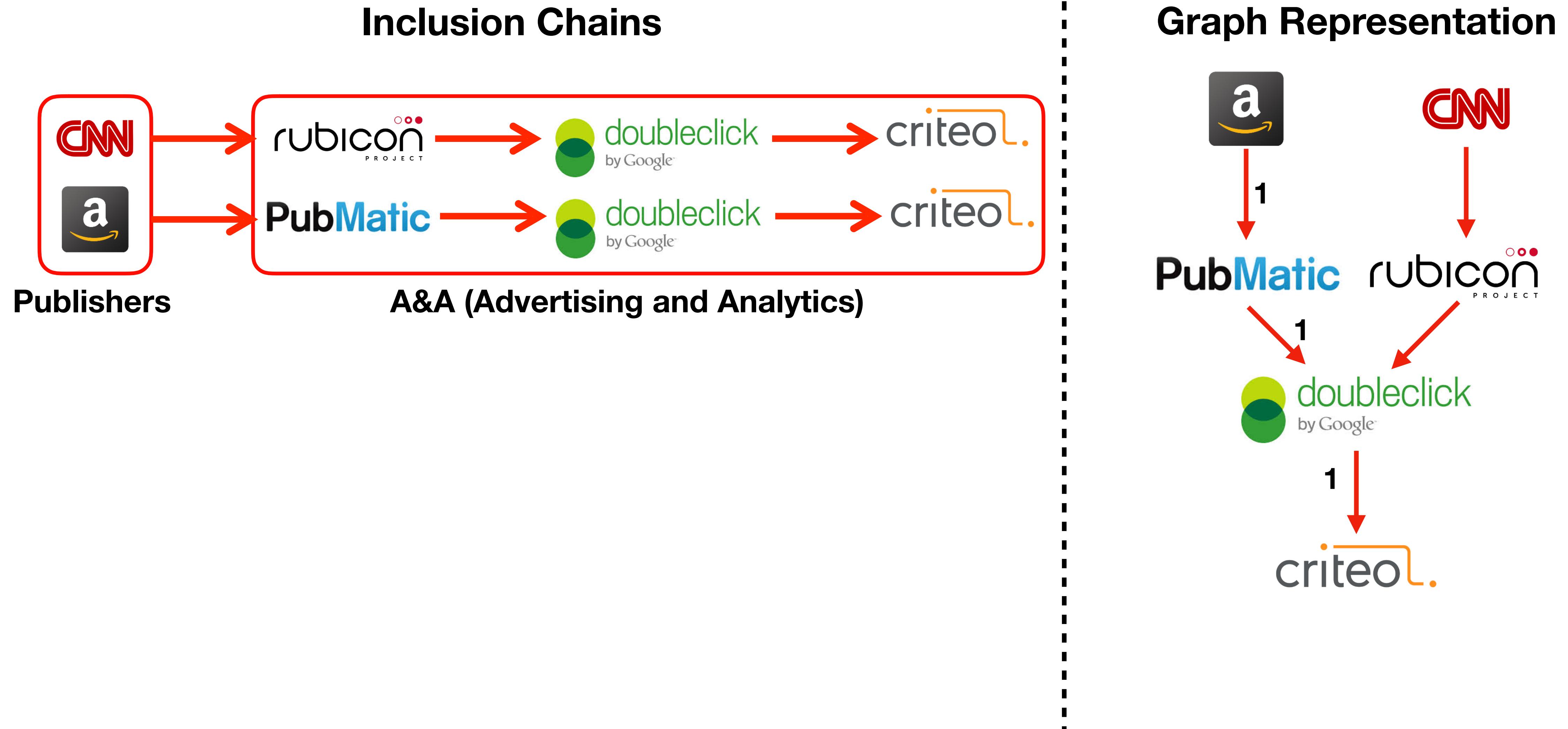
From Chains to Graph



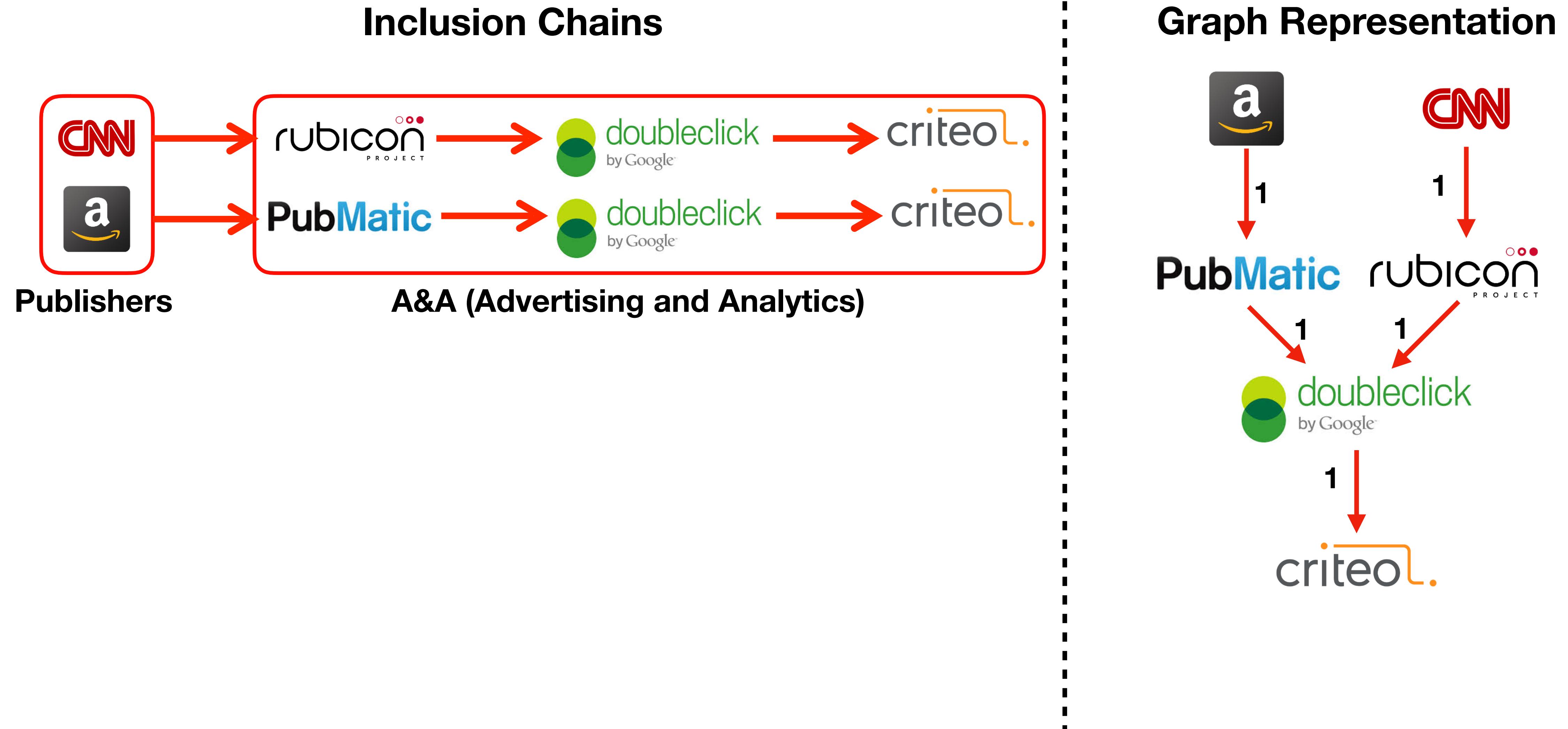
From Chains to Graph



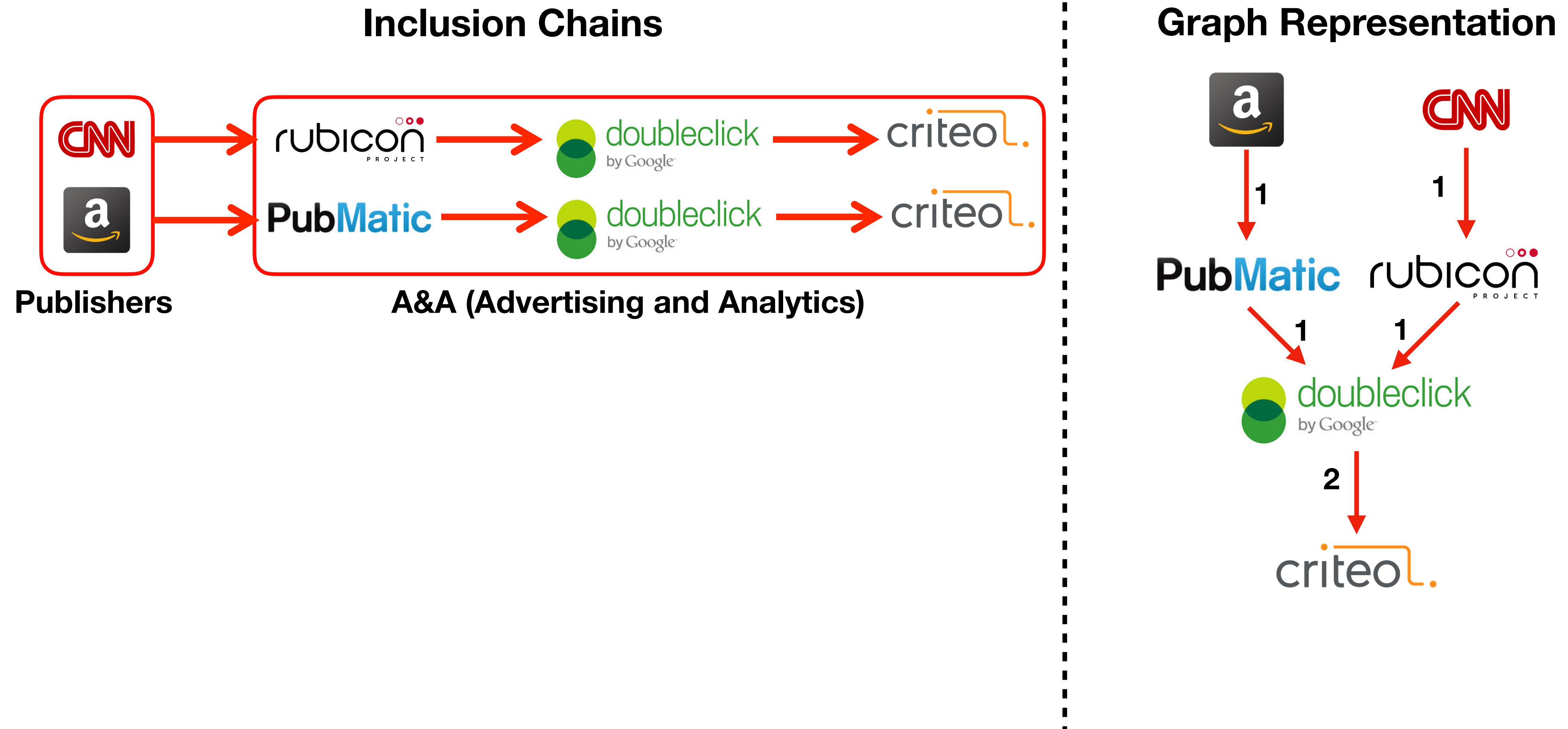
From Chains to Graph



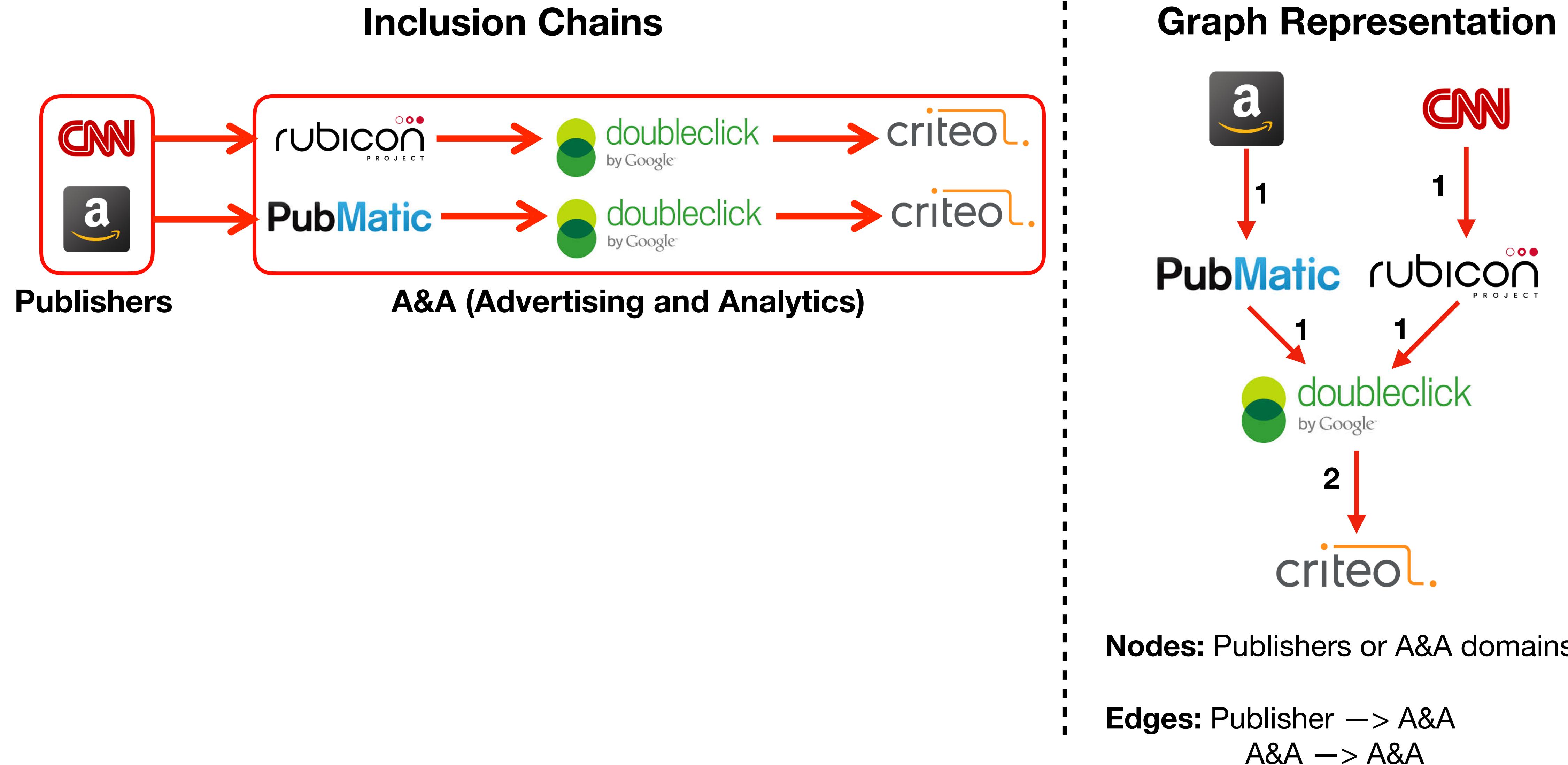
From Chains to Graph



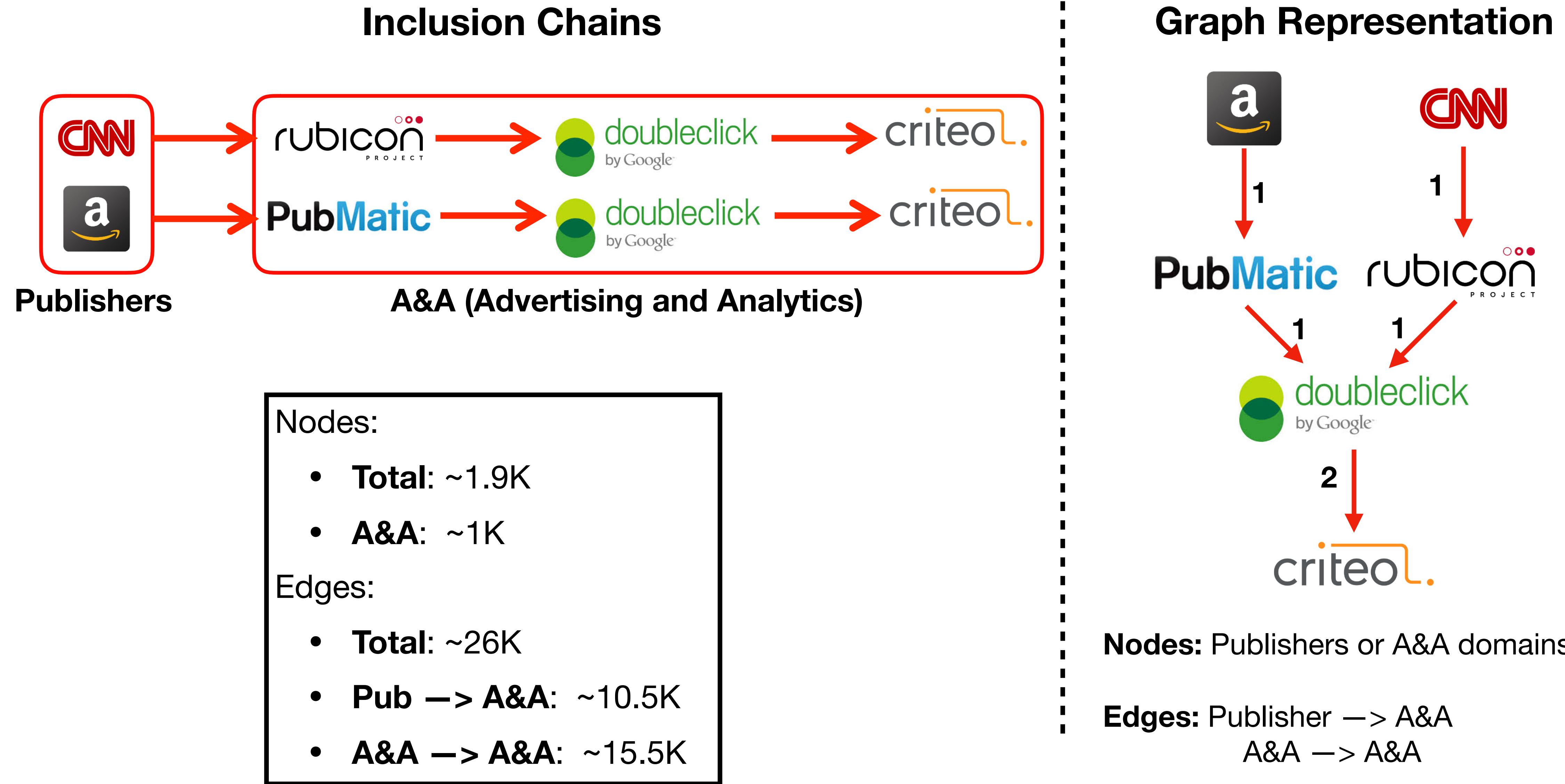
From Chains to Graph



From Chains to Graph



From Chains to Graph



Simulation Setup for Propagation of Tracking Data

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:
 - A. **Direct Propagation:** 
 - Present on publisher or won RTB auction. **Observable** (goes through the browser)

[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

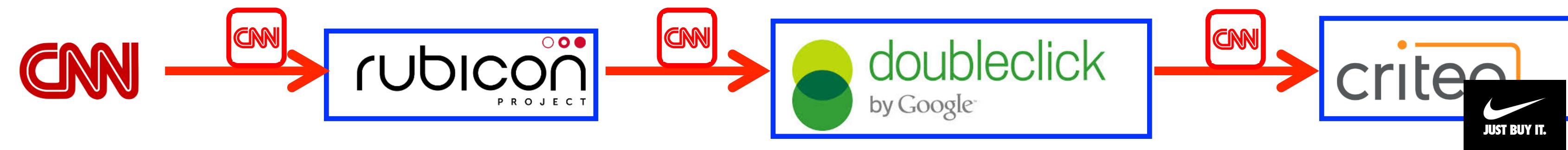
Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:

A. **Direct Propagation:**

- Present on publisher or won RTB auction. **Observable** (goes through the browser)



[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

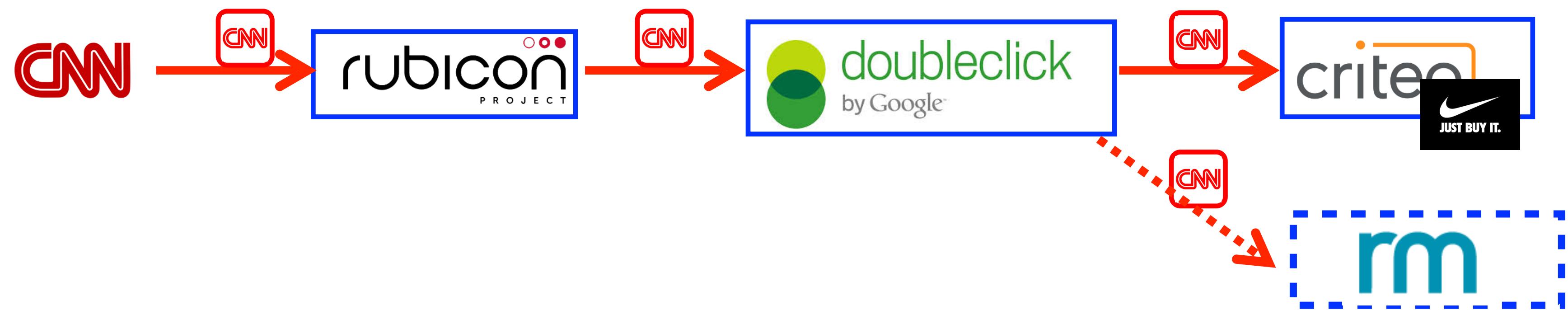
1. User generates an impression on N selected publishers.
2. Impressions are forwarded to A&A domains via:

A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

- A&A domains learn impressions through RTB participation. **Non-observable**



[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Setup for Propagation of Tracking Data

We simulate browsing traces for 200 users using method from [1].

1. User generates an impression on N selected publishers.

2. Impressions are forwarded to A&A domains via:

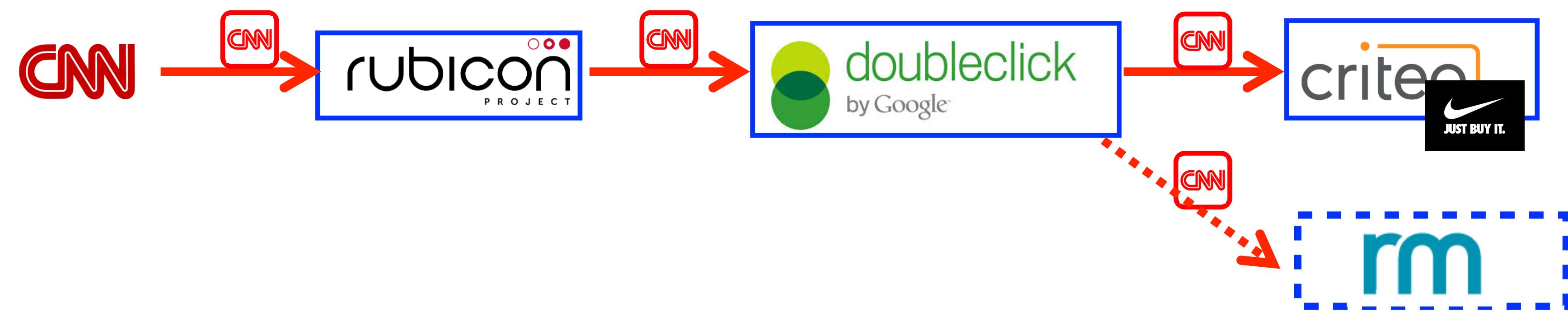
A. **Direct Propagation:** 

- Present on publisher or won RTB auction. **Observable** (goes through the browser)

B. **Indirect Propagation:** 

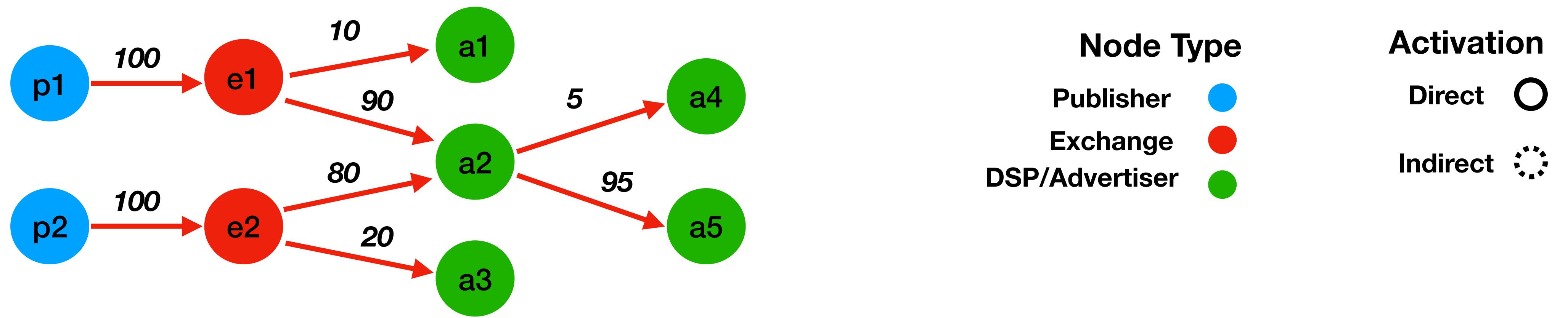
- A&A domains learn impressions through RTB participation. **Non-observable**

3. RTB winner is decided based on probability (function of edge weights).

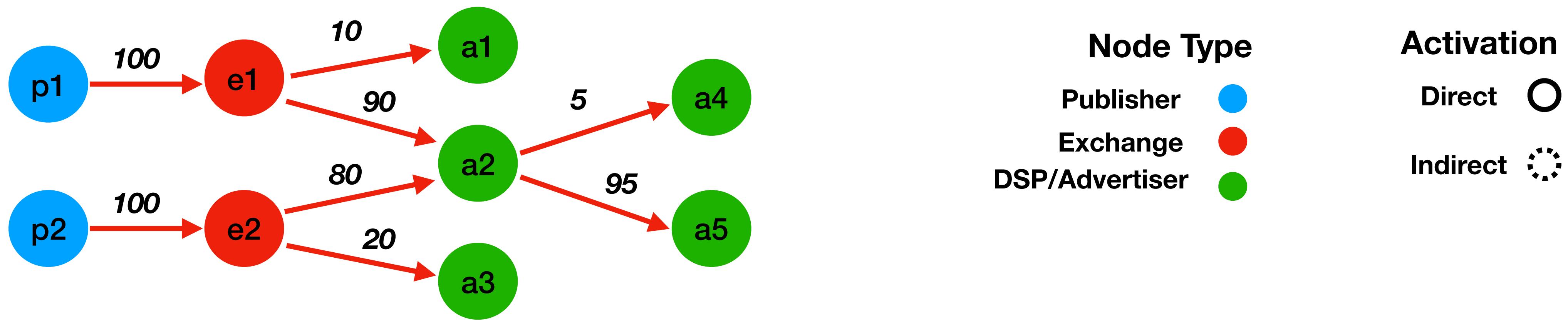


[1]. Burken et al. User centric walk: An integrated approach for modeling the browsing behavior of users on the web. ASS 2005

Simulation Example

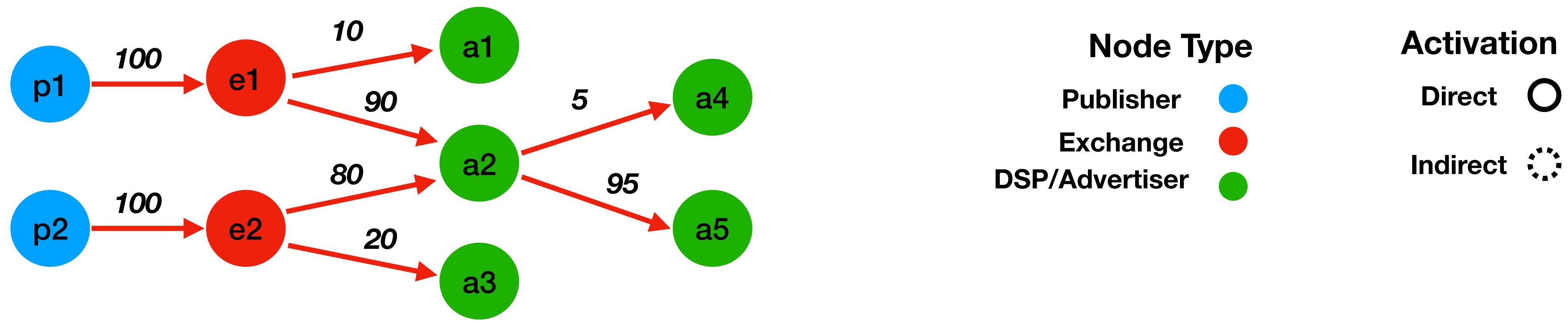


Simulation Example



We use the list of A&A domains
isolated from ads.txt as
Ad Exchanges

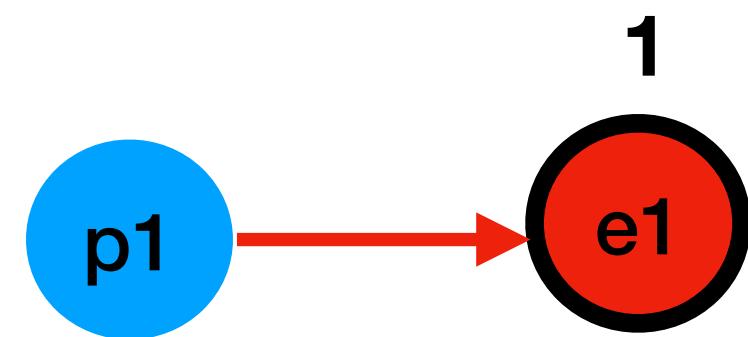
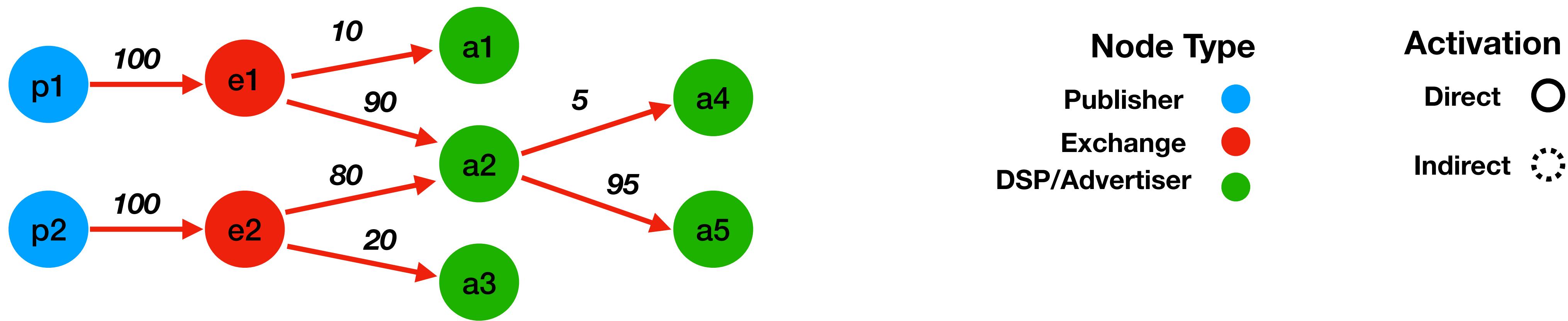
Simulation Example



p1

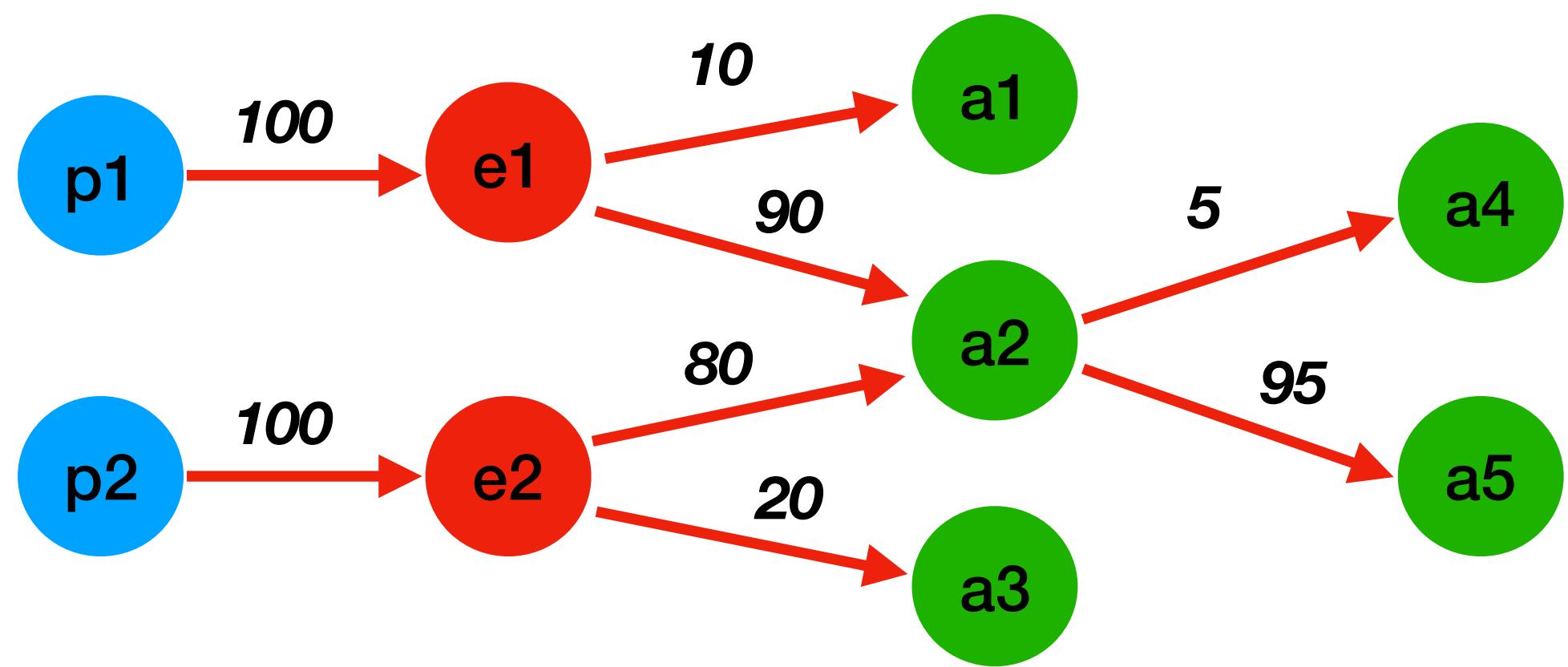
We use the list of A&A domains
isolated from ads.txt as
Ad Exchanges

Simulation Example

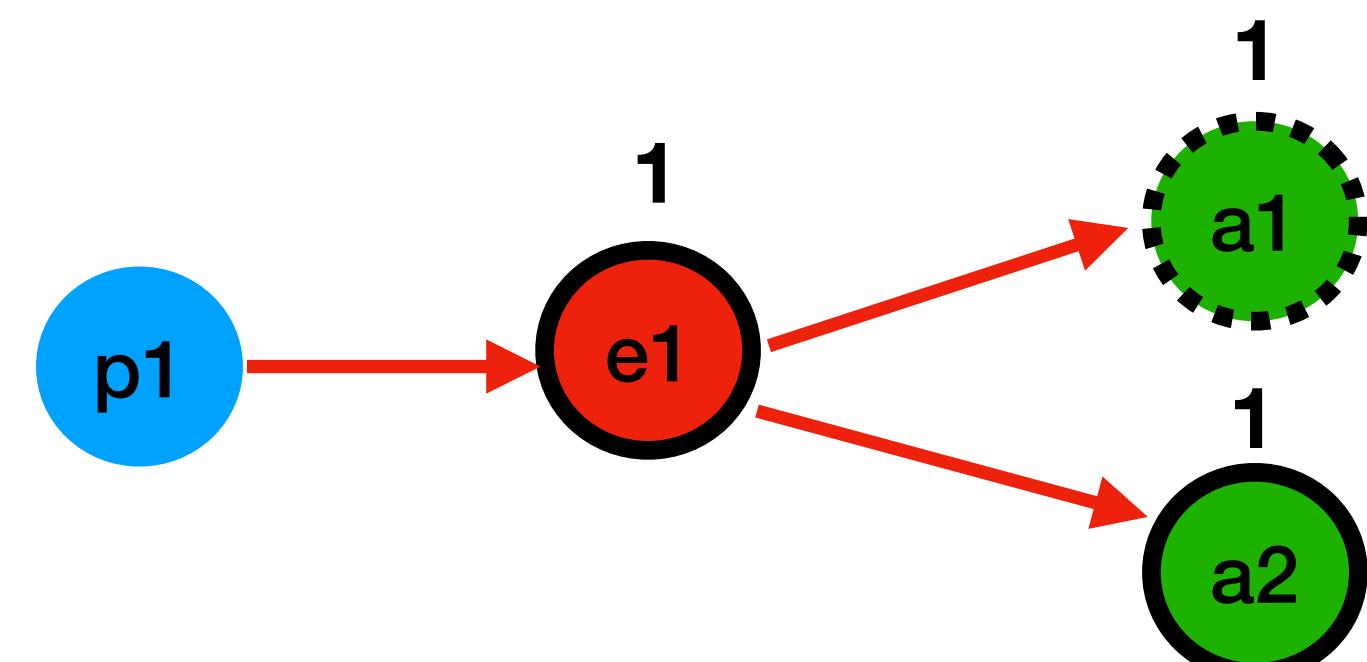


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example

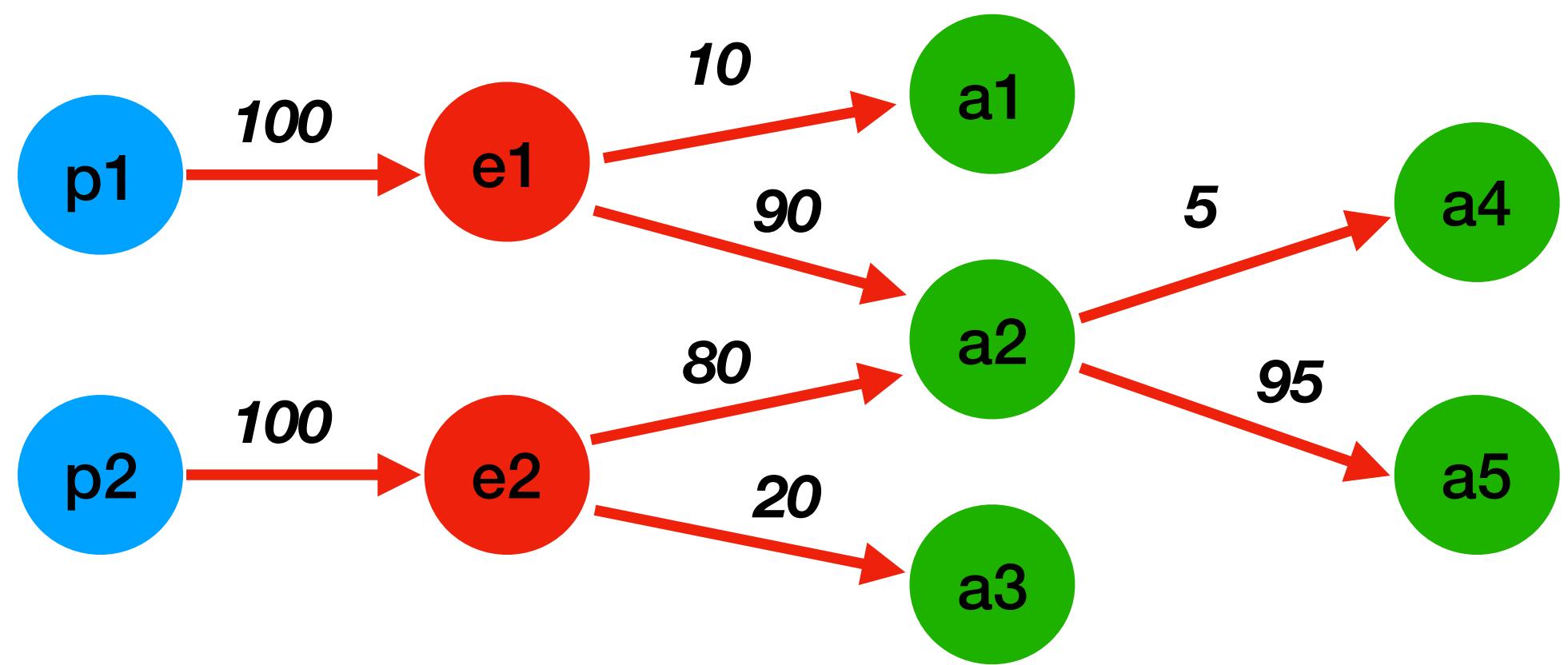


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect

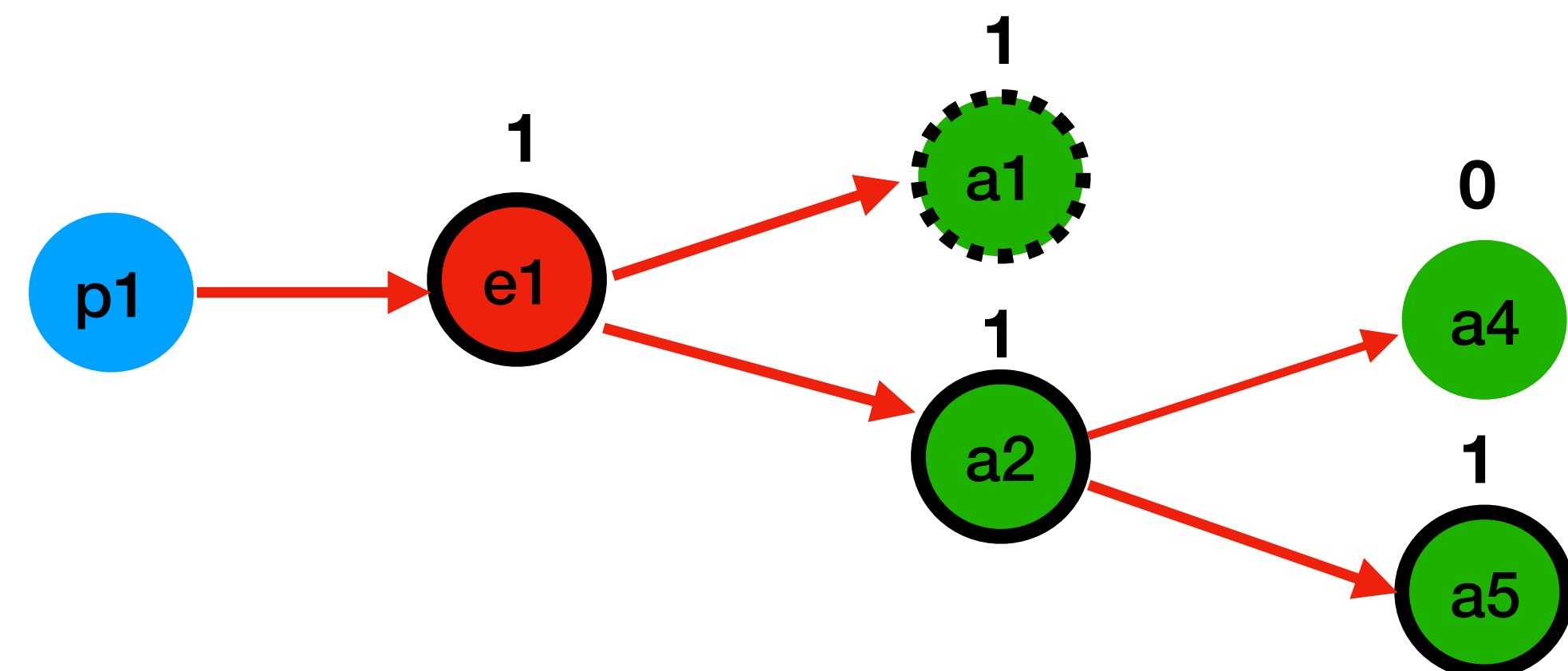


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example

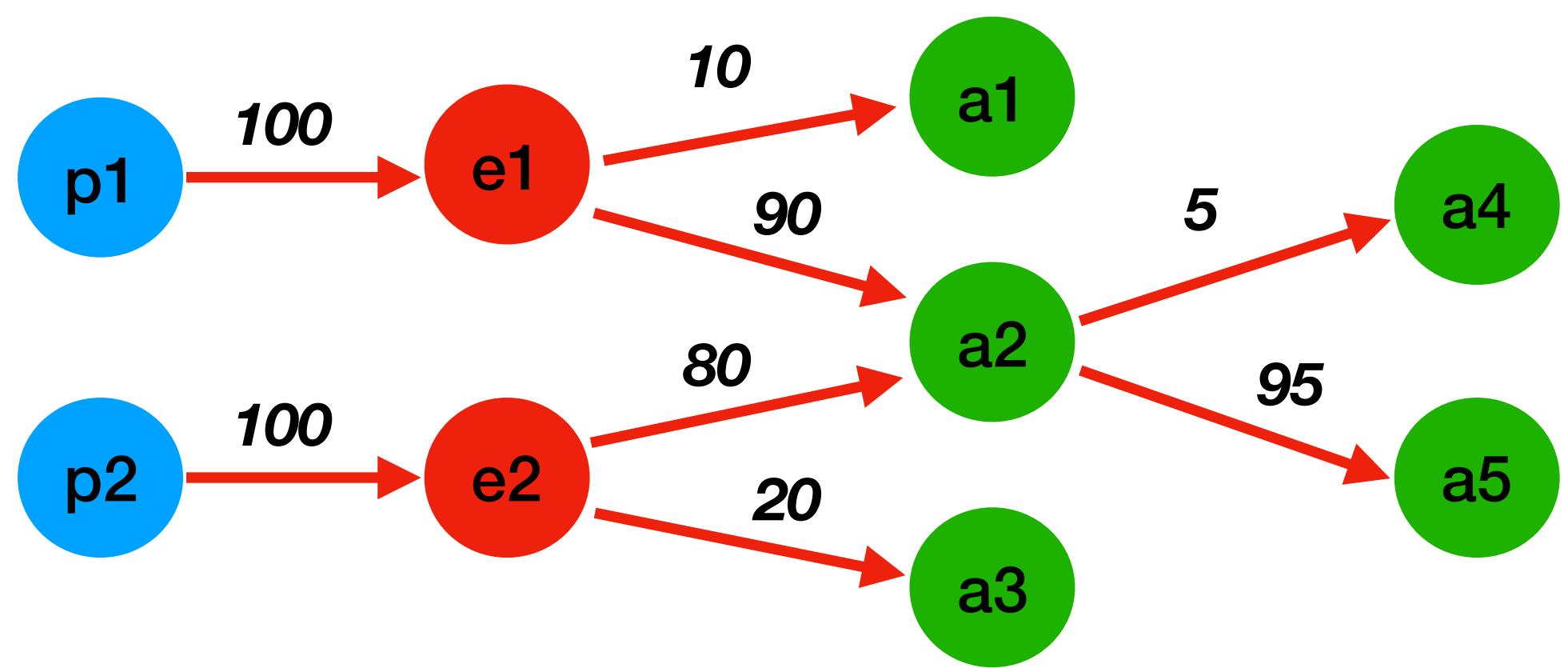


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect

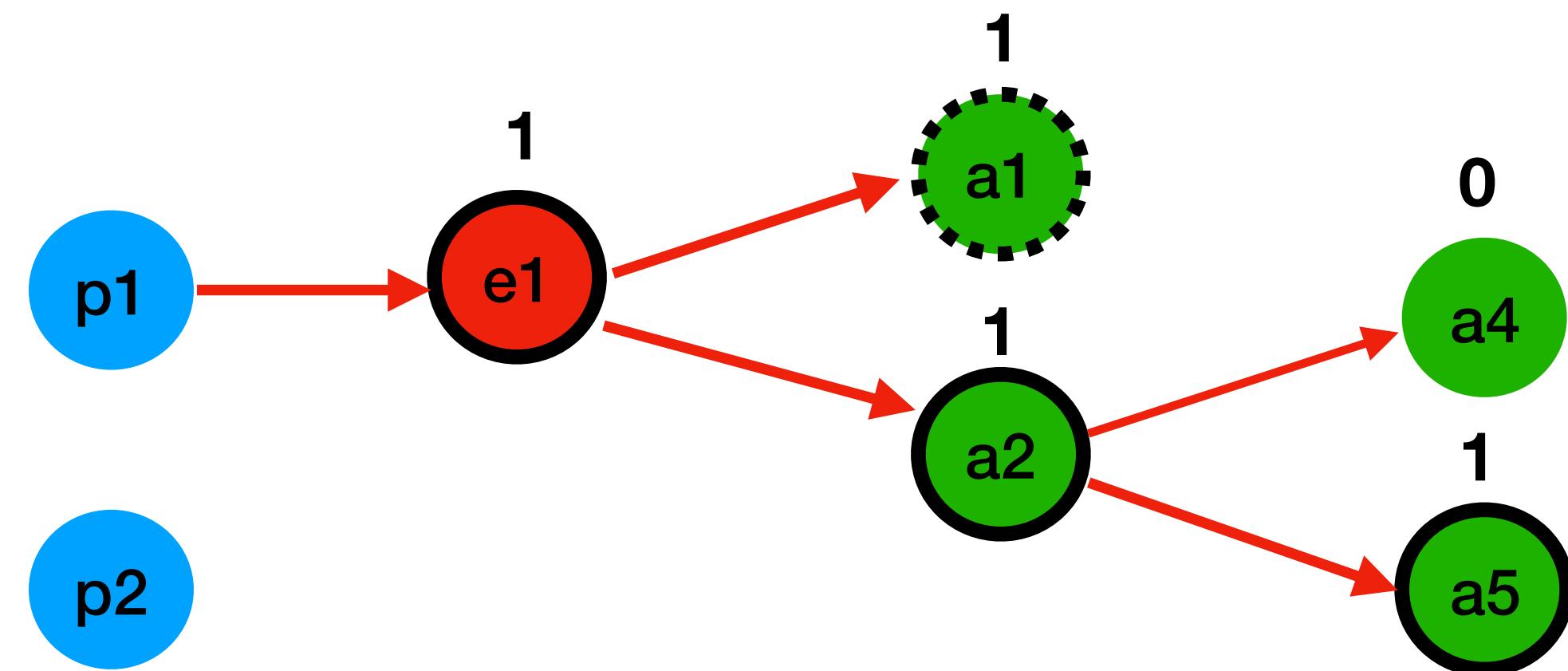


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example

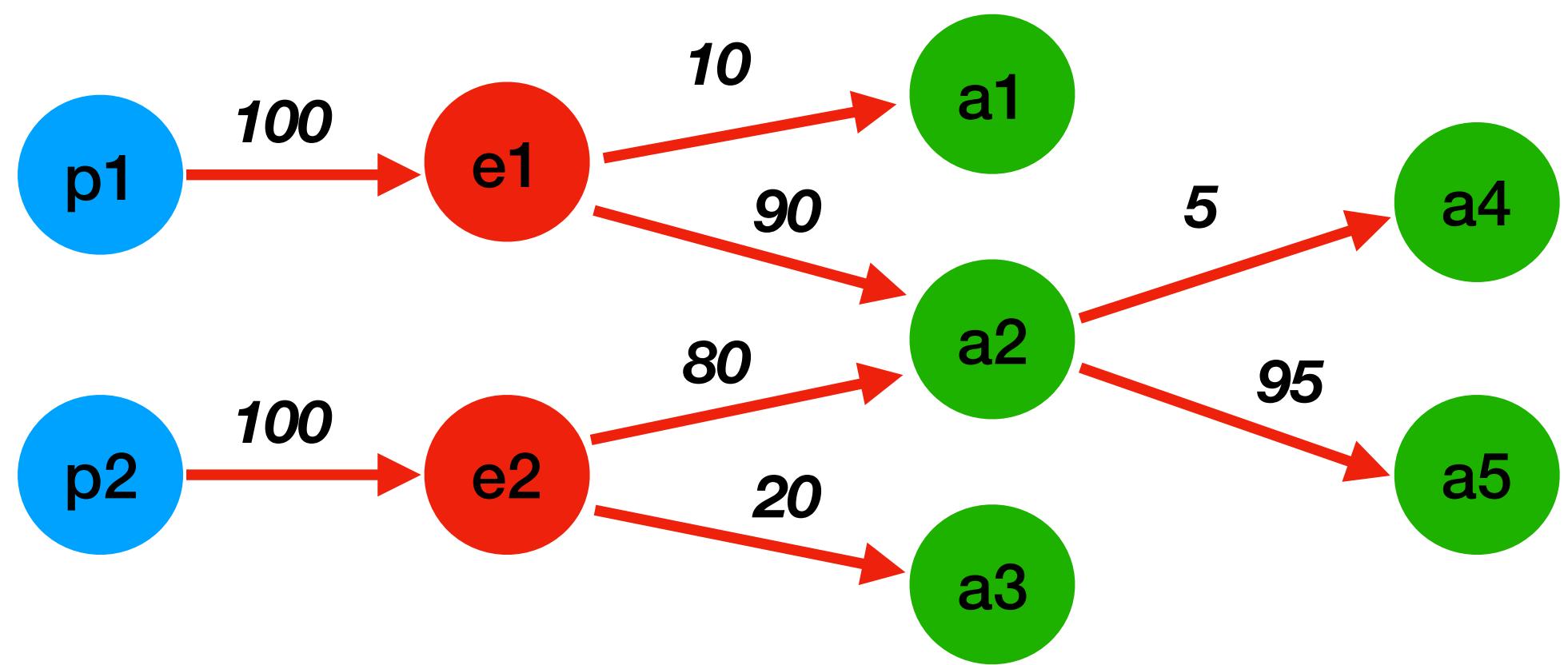


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect

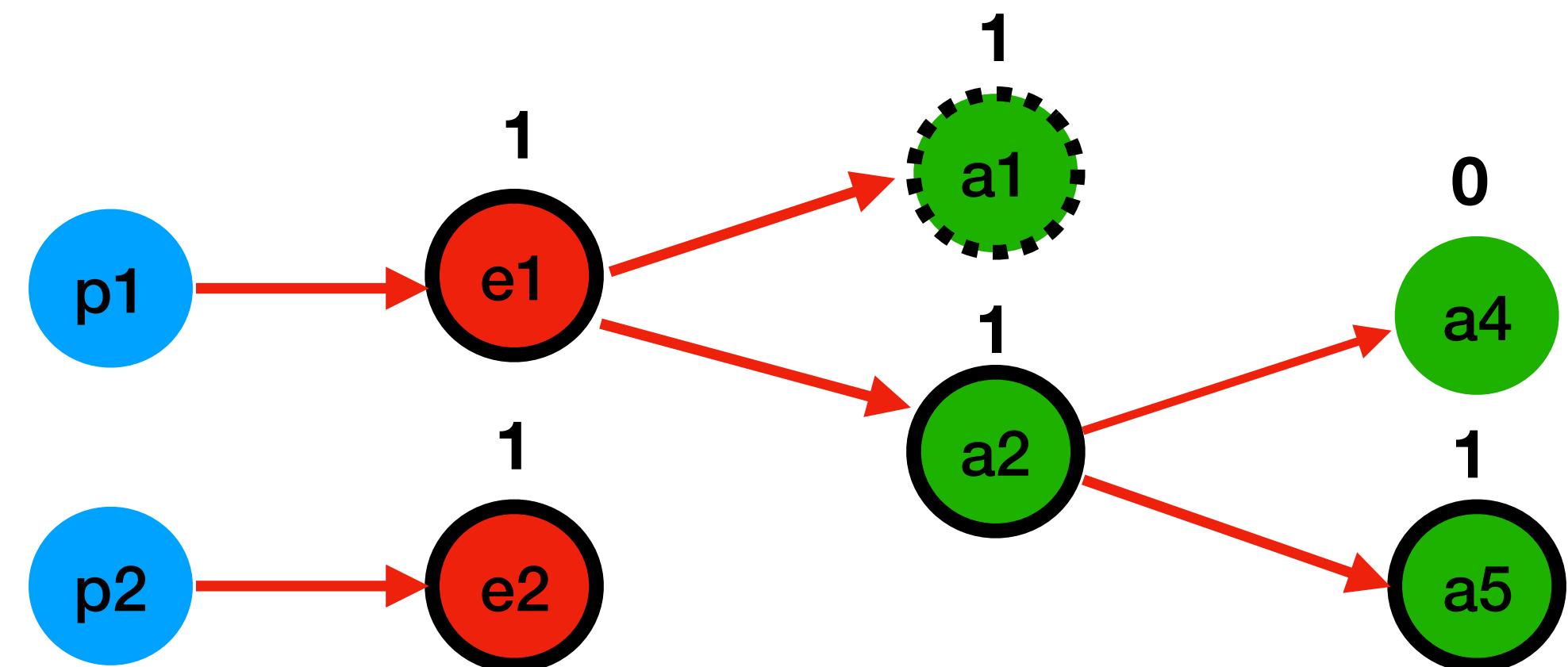


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example

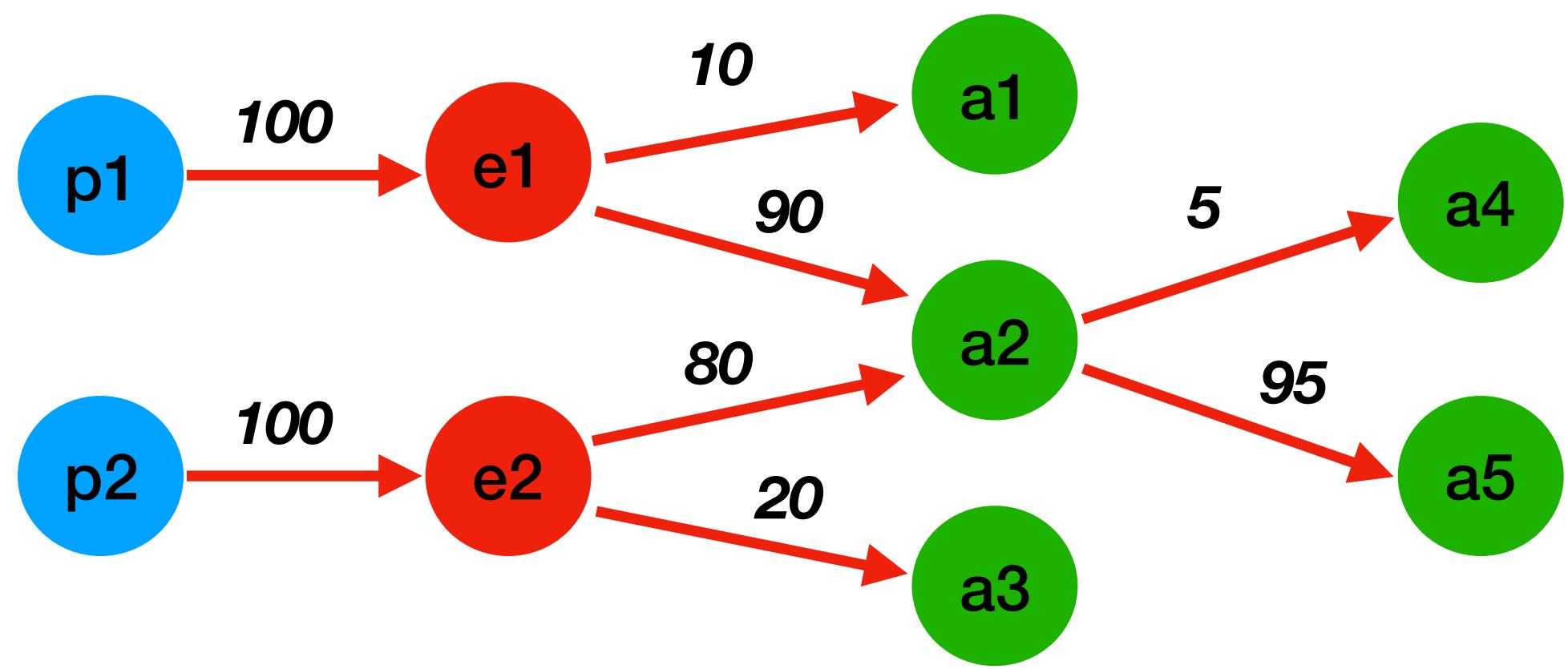


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect

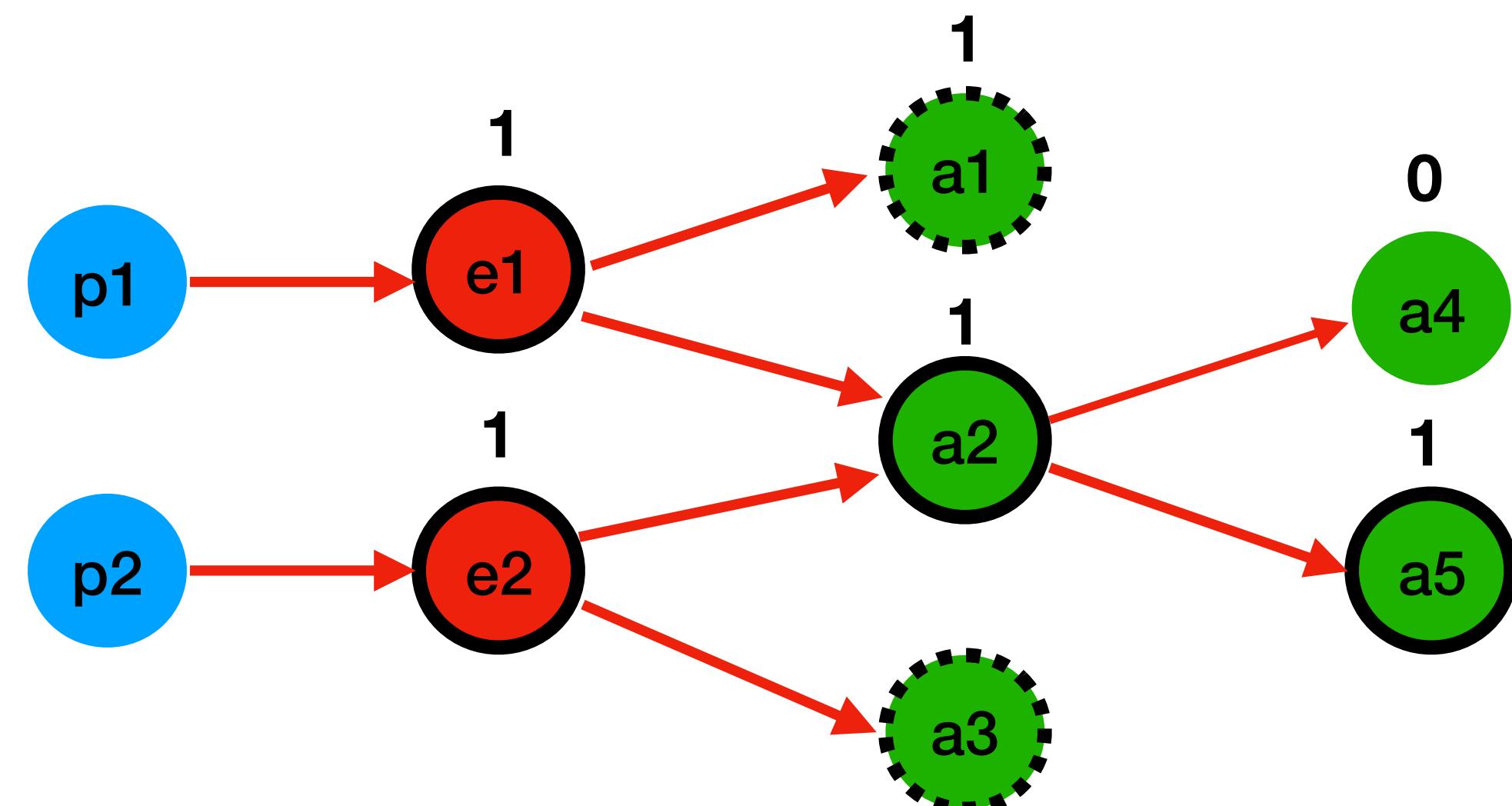


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example

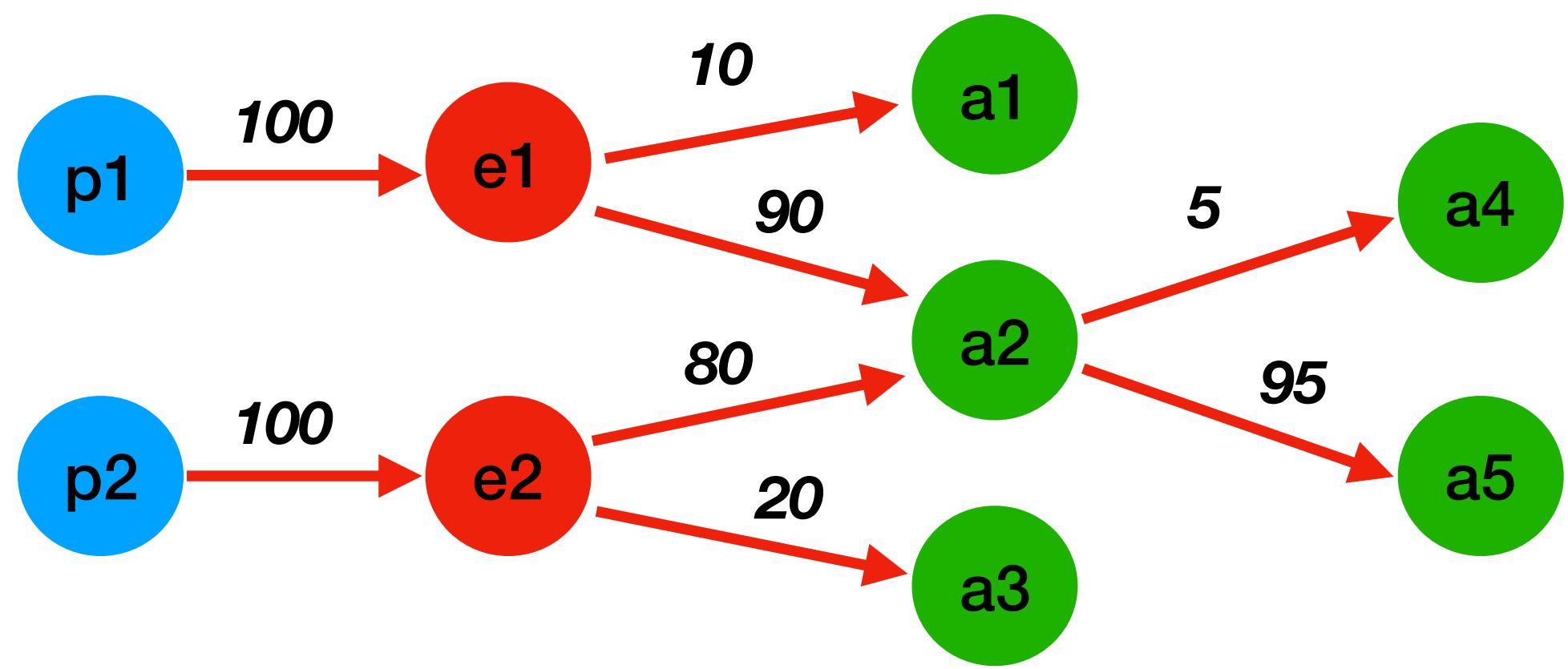


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect

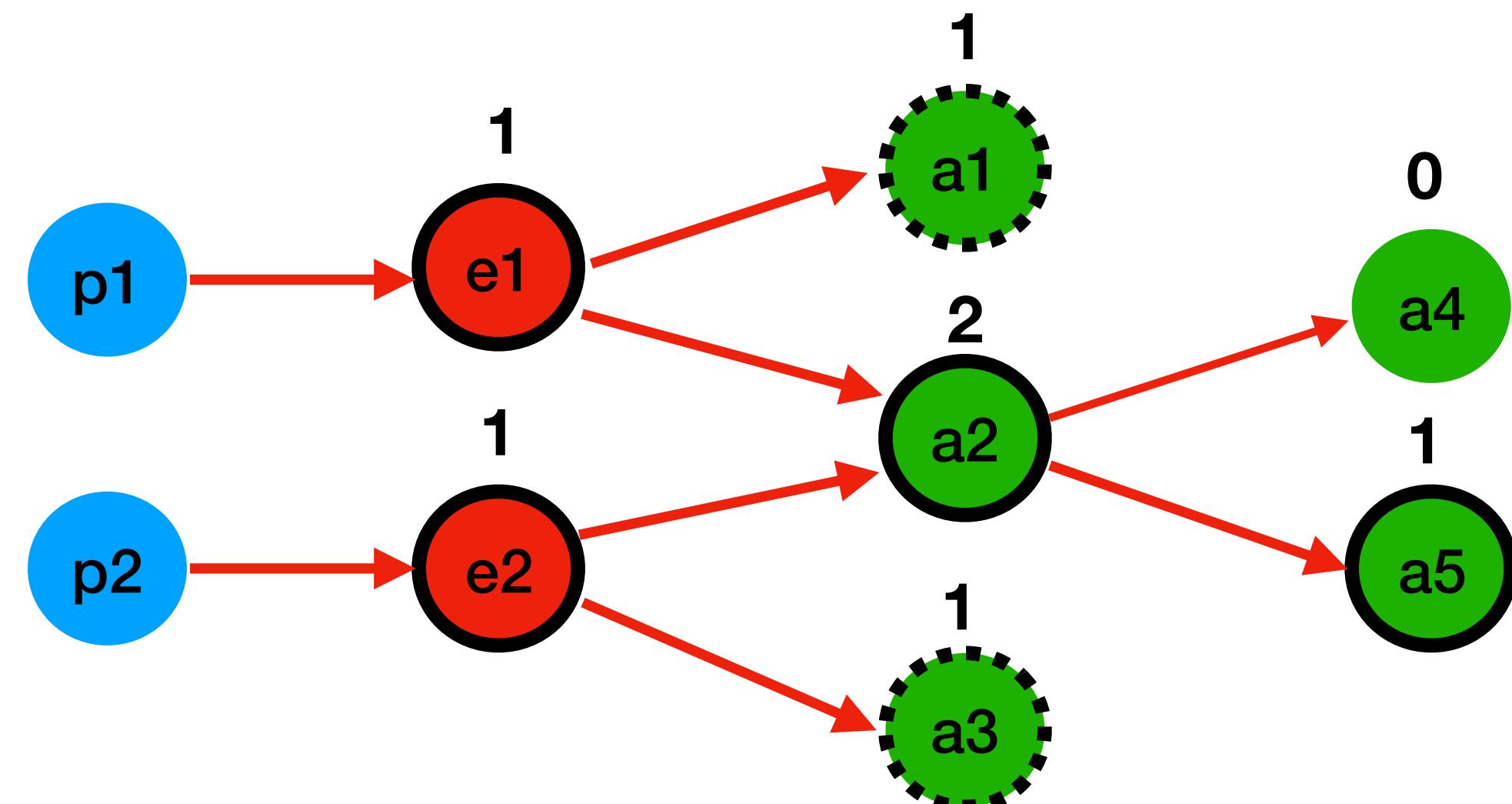


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example

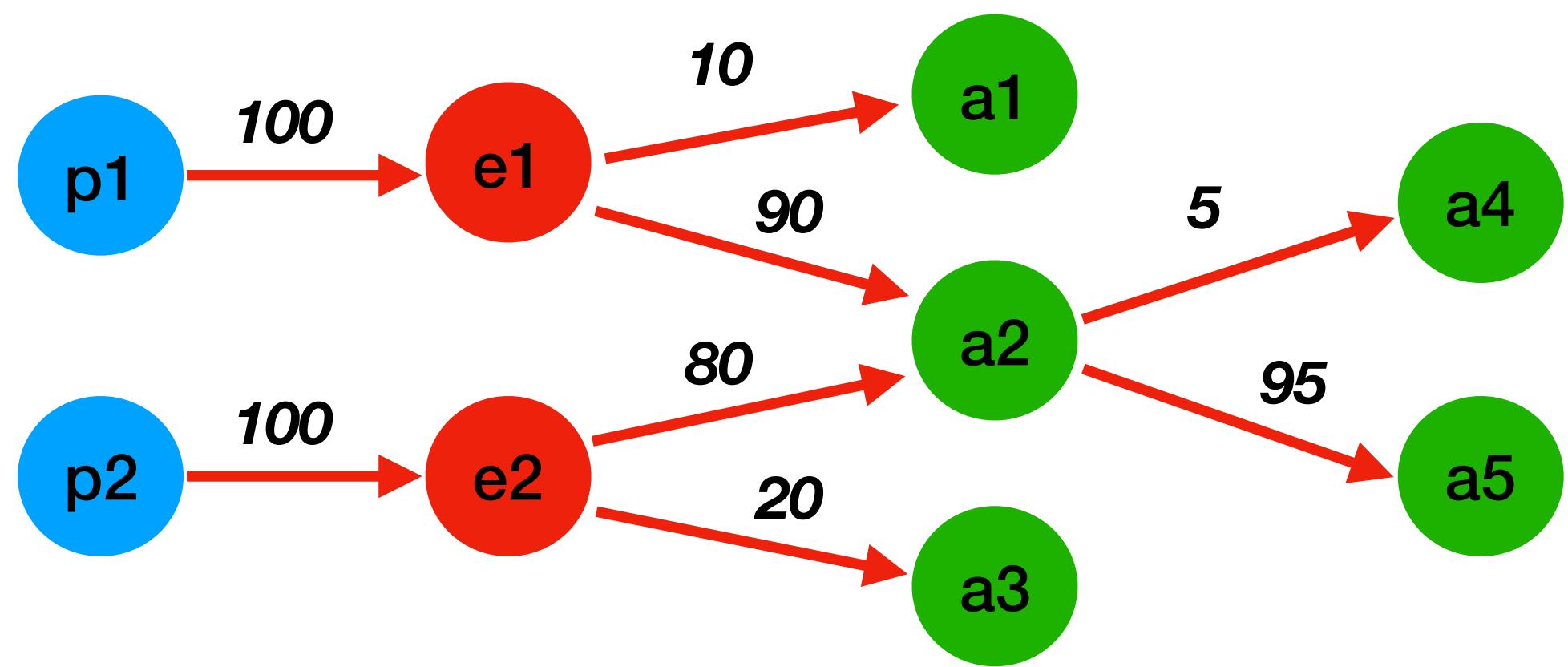


Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect

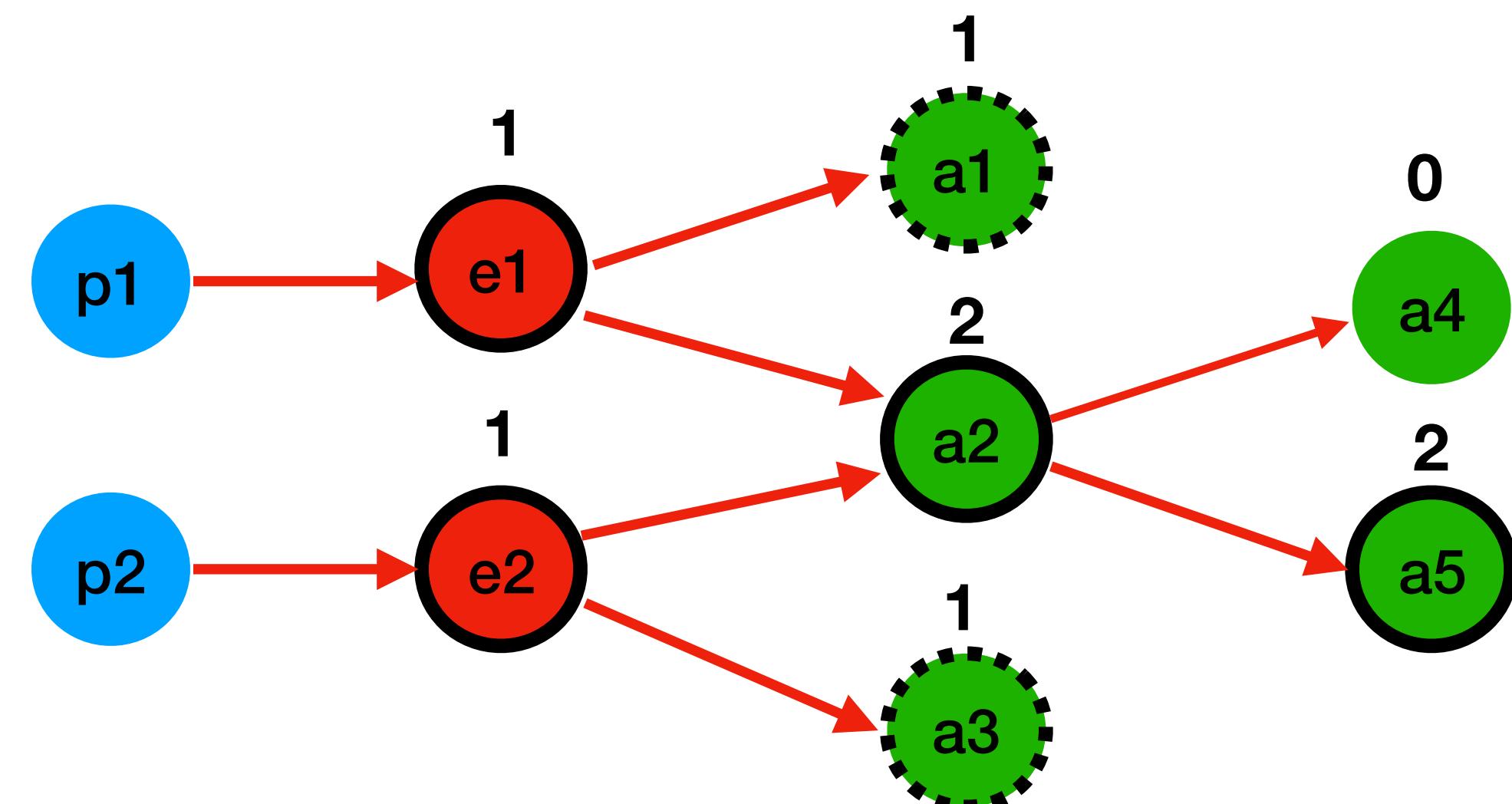


We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Simulation Example



Node Type	Activation
Publisher	Direct
Exchange	Indirect
DSP/Advertiser	Indirect



We use the list of A&A domains isolated from ads.txt as Ad Exchanges

Impressions Observed

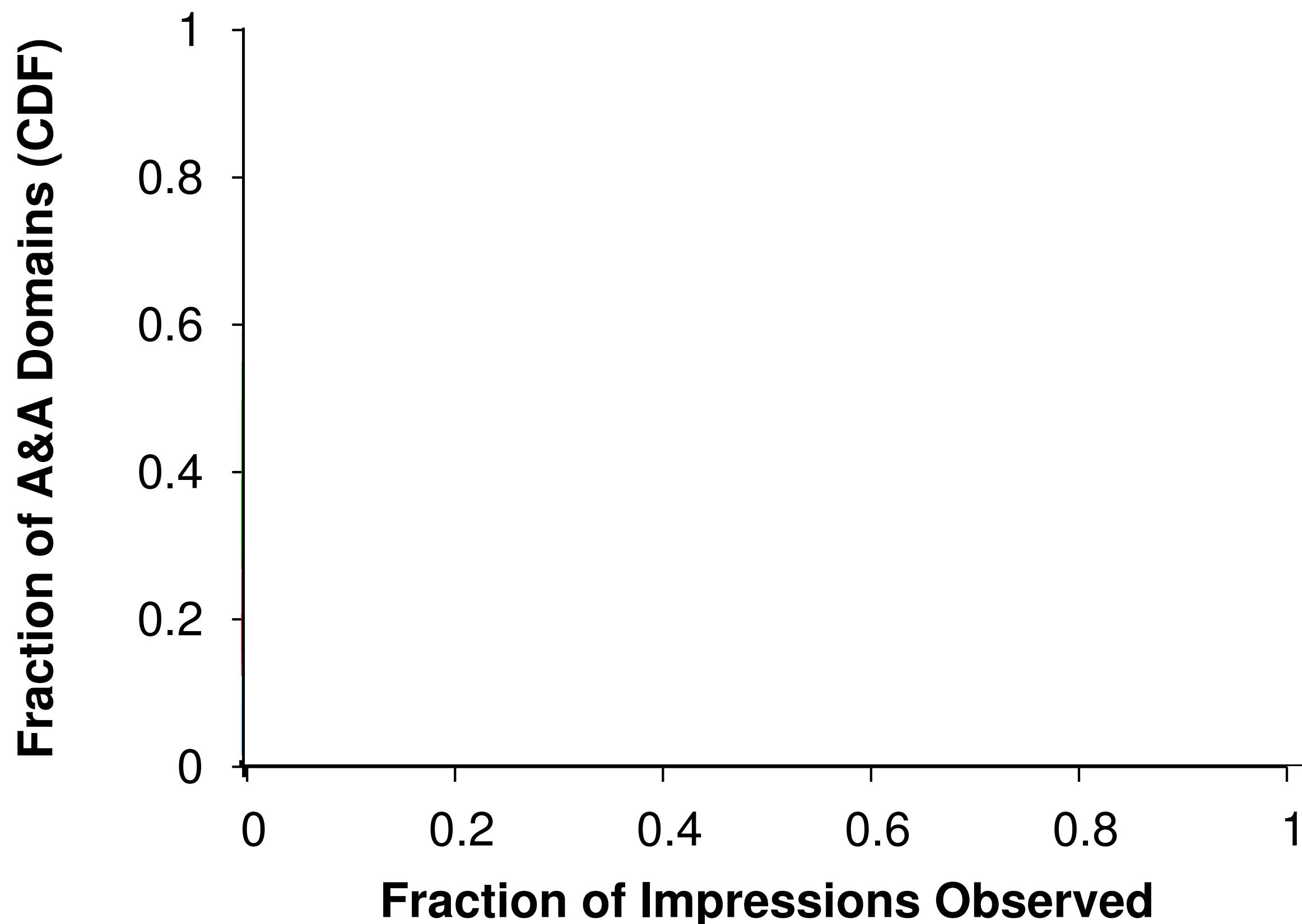
We have 3 simulation models:

1. **RTB-Relaxed:** Upper-bound
 - All A&A domains act as ad exchanges
2. **Cookie-Matching:** Lower-bound
 - Indirect impressions forwarded along 200 edges only (No ad exchanges).
3. **RTB-Constrained:** Realistic Scenario
 - Ad exchanges selected from ads.txt data.

Impressions Observed

We have 3 simulation models:

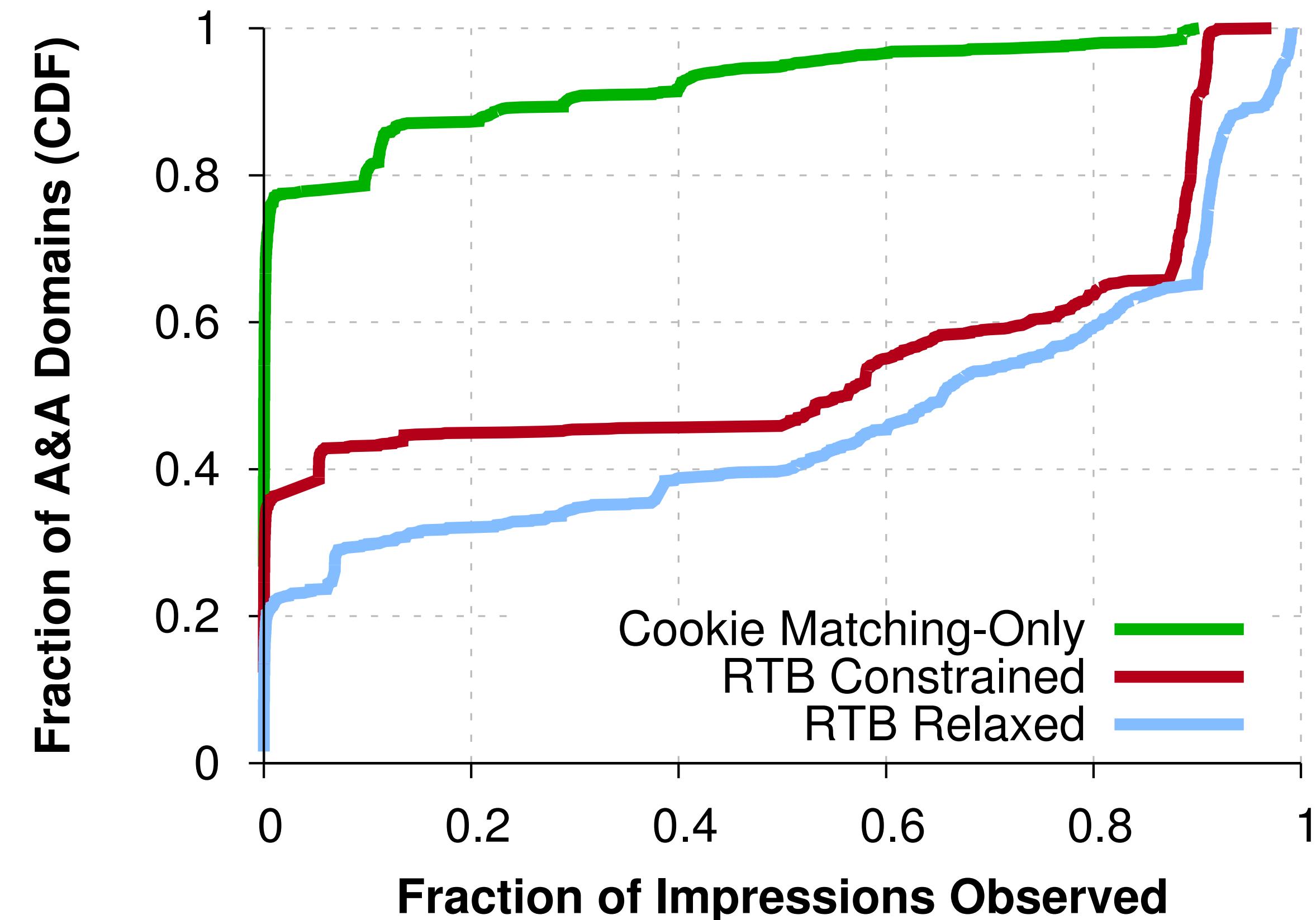
1. **RTB-Relaxed:** Upper-bound
 - All A&A domains act as ad exchanges
2. **Cookie-Matching:** Lower-bound
 - Indirect impressions forwarded along 200 edges only (No ad exchanges).
3. **RTB-Constrained:** Realistic Scenario
 - Ad exchanges selected from ads.txt data.



Impressions Observed

We have 3 simulation models:

1. **RTB-Relaxed:** Upper-bound
 - All A&A domains act as ad exchanges
2. **Cookie-Matching:** Lower-bound
 - Indirect impressions forwarded along 200 edges only (No ad exchanges).
3. **RTB-Constrained:** Realistic Scenario
 - Ad exchanges selected from ads.txt data.



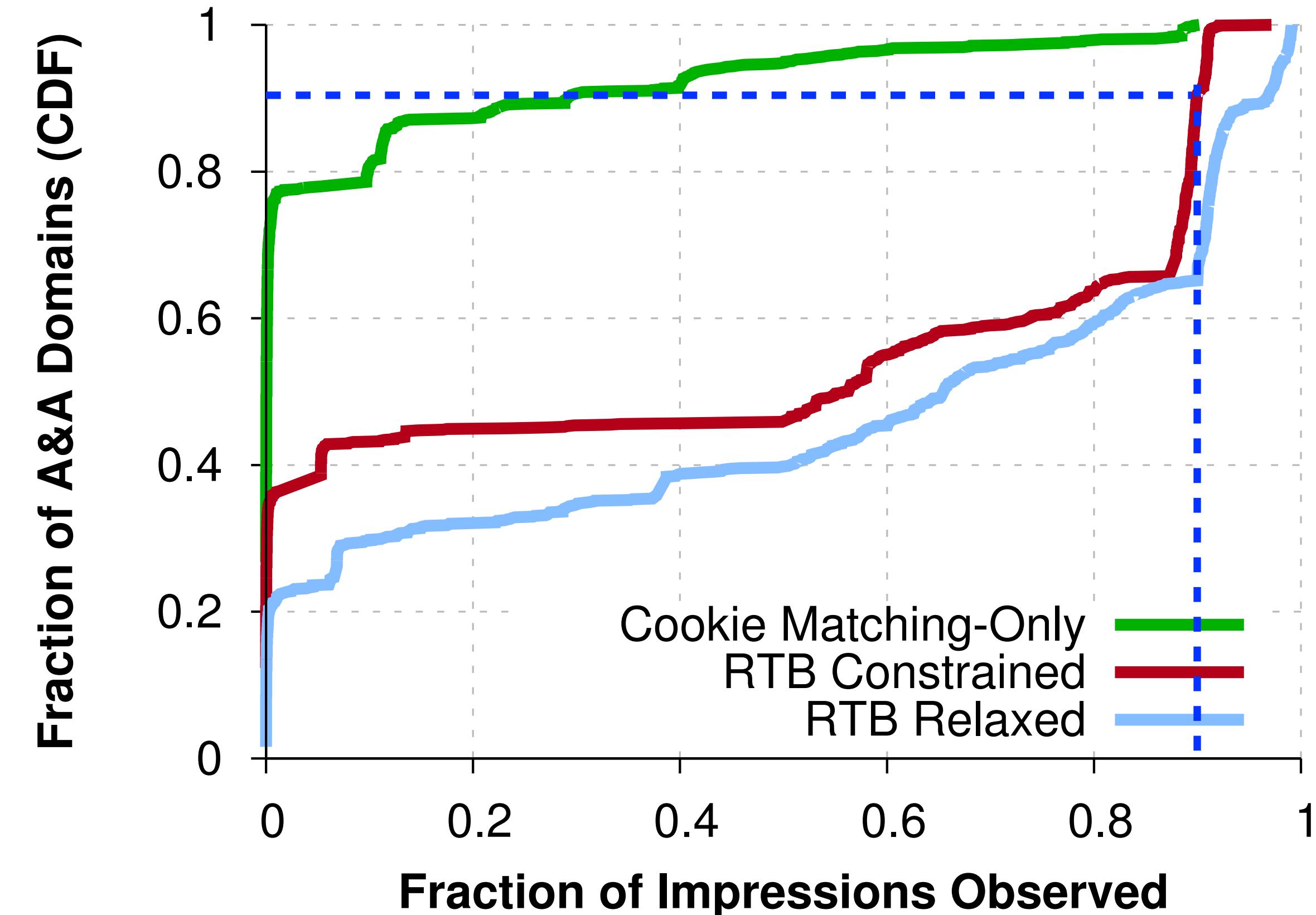
Impressions Observed

We have 3 simulation models:

1. **RTB-Relaxed:** Upper-bound
 - All A&A domains act as ad exchanges
2. **Cookie-Matching:** Lower-bound
 - Indirect impressions forwarded along 200 edges only (No ad exchanges).
3. **RTB-Constrained:** Realistic Scenario
 - Ad exchanges selected from ads.txt data.

Take Aways

1. RTB-Constrained is very close to RTB-Relaxed
2. 10% A&A see more than 90% of impressions in RTB-Constrained



Effect of Ad Exchanges

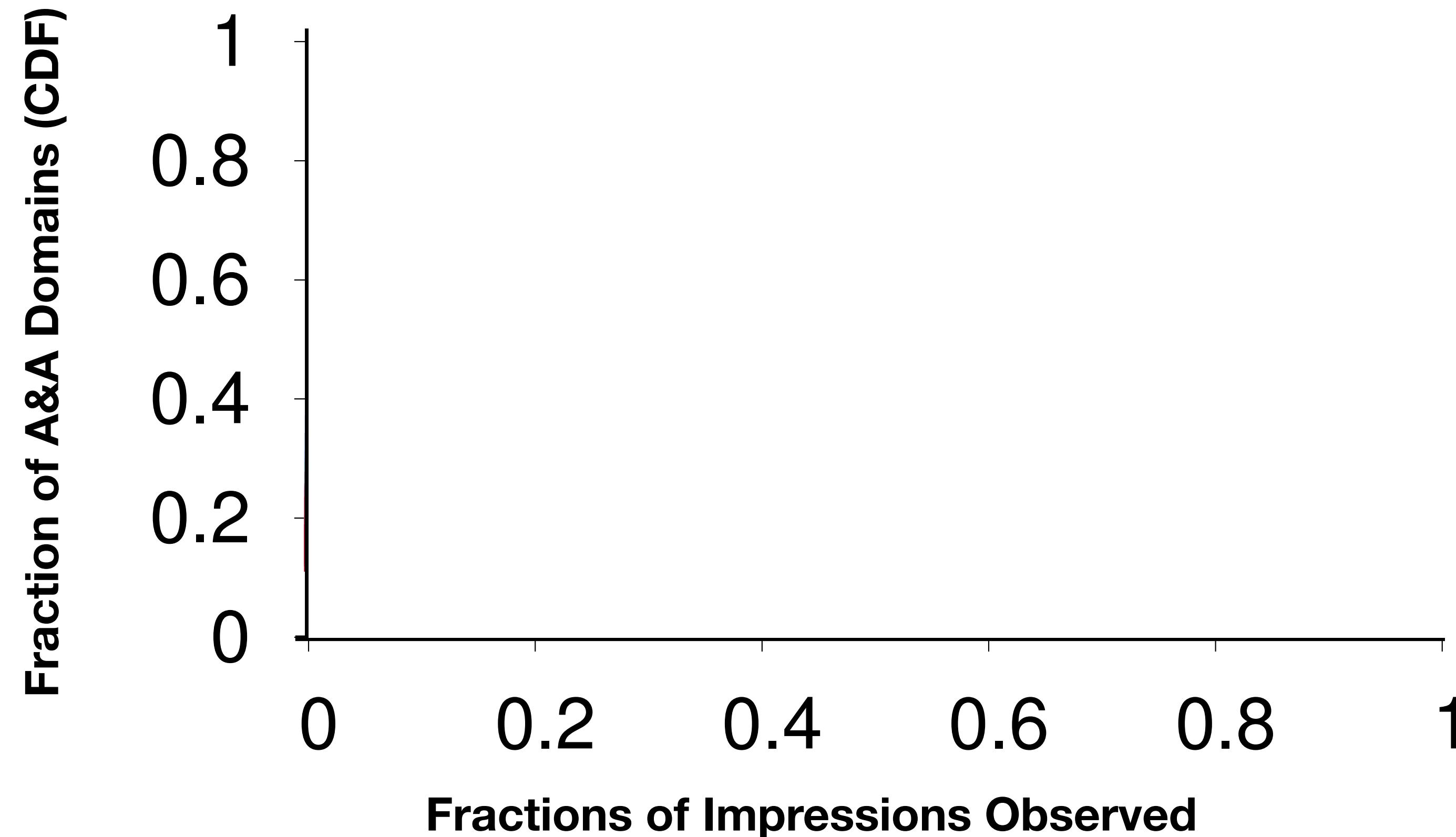


Fig: Fraction of impressions observed by A&A domains under RTB-C model when top x exchanges are selected

Effect of Ad Exchanges

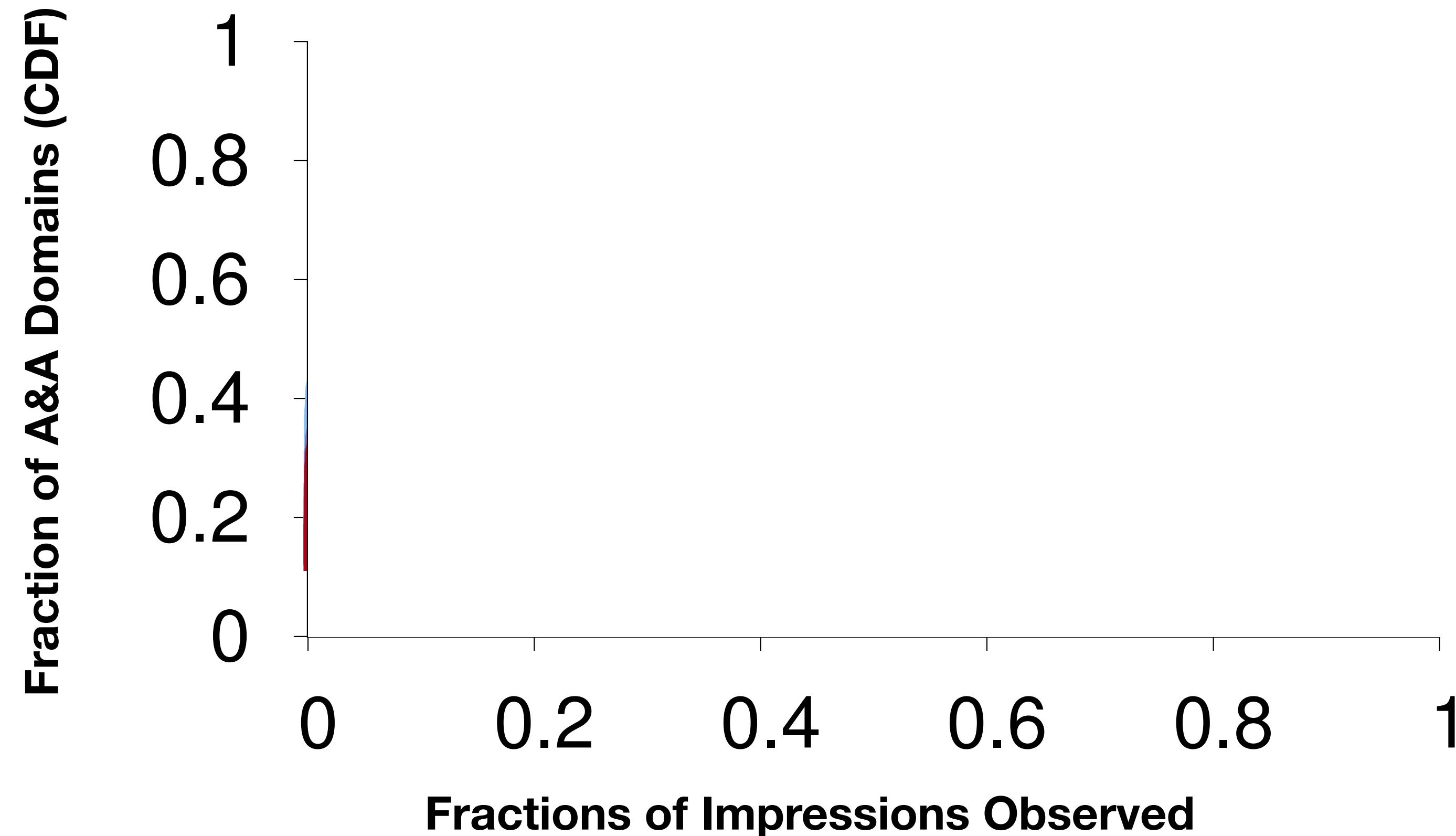


Fig: Fraction of impressions observed by A&A domains under RTB-C model when top x exchanges are selected

Effect of Ad Exchanges

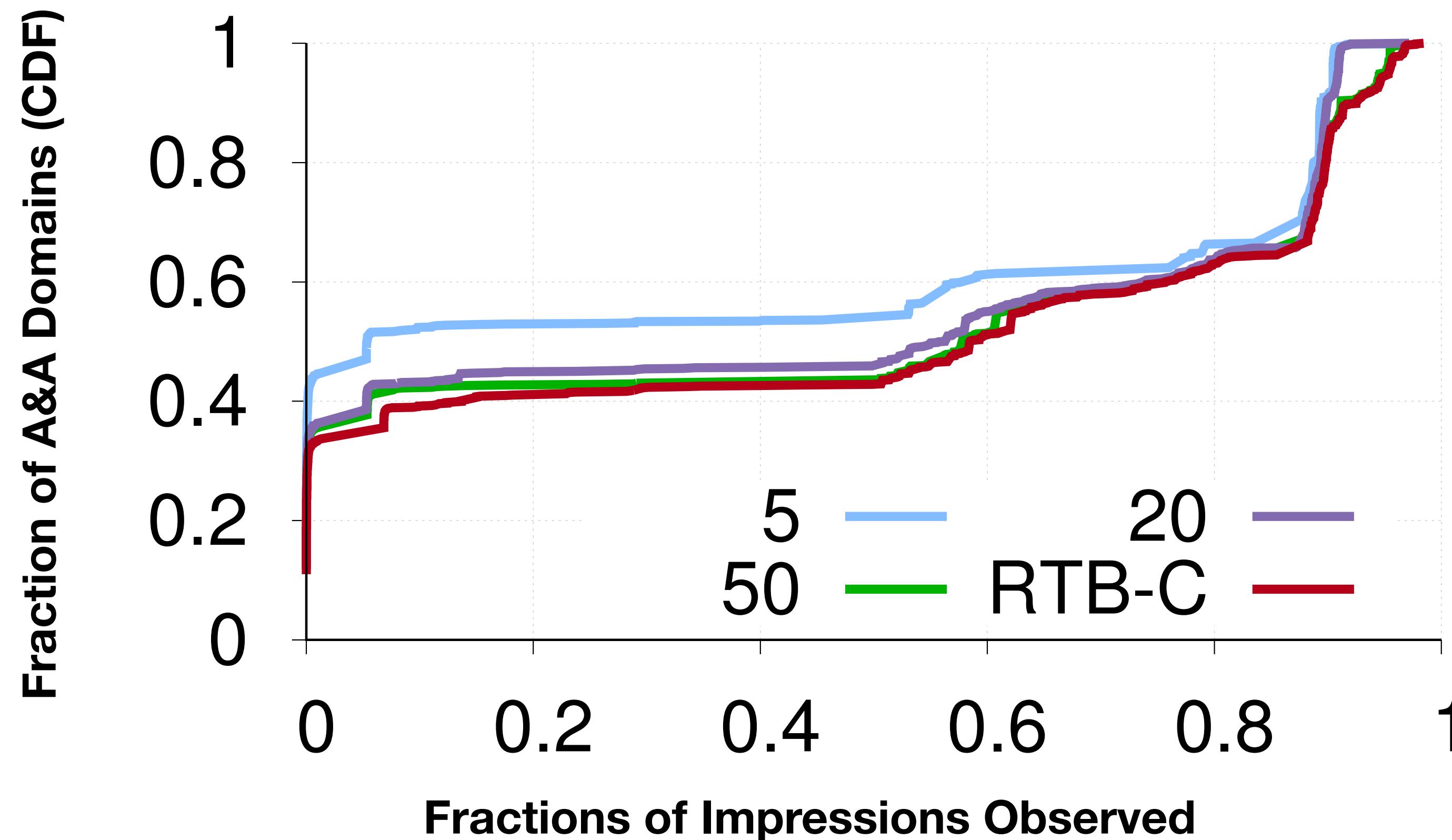
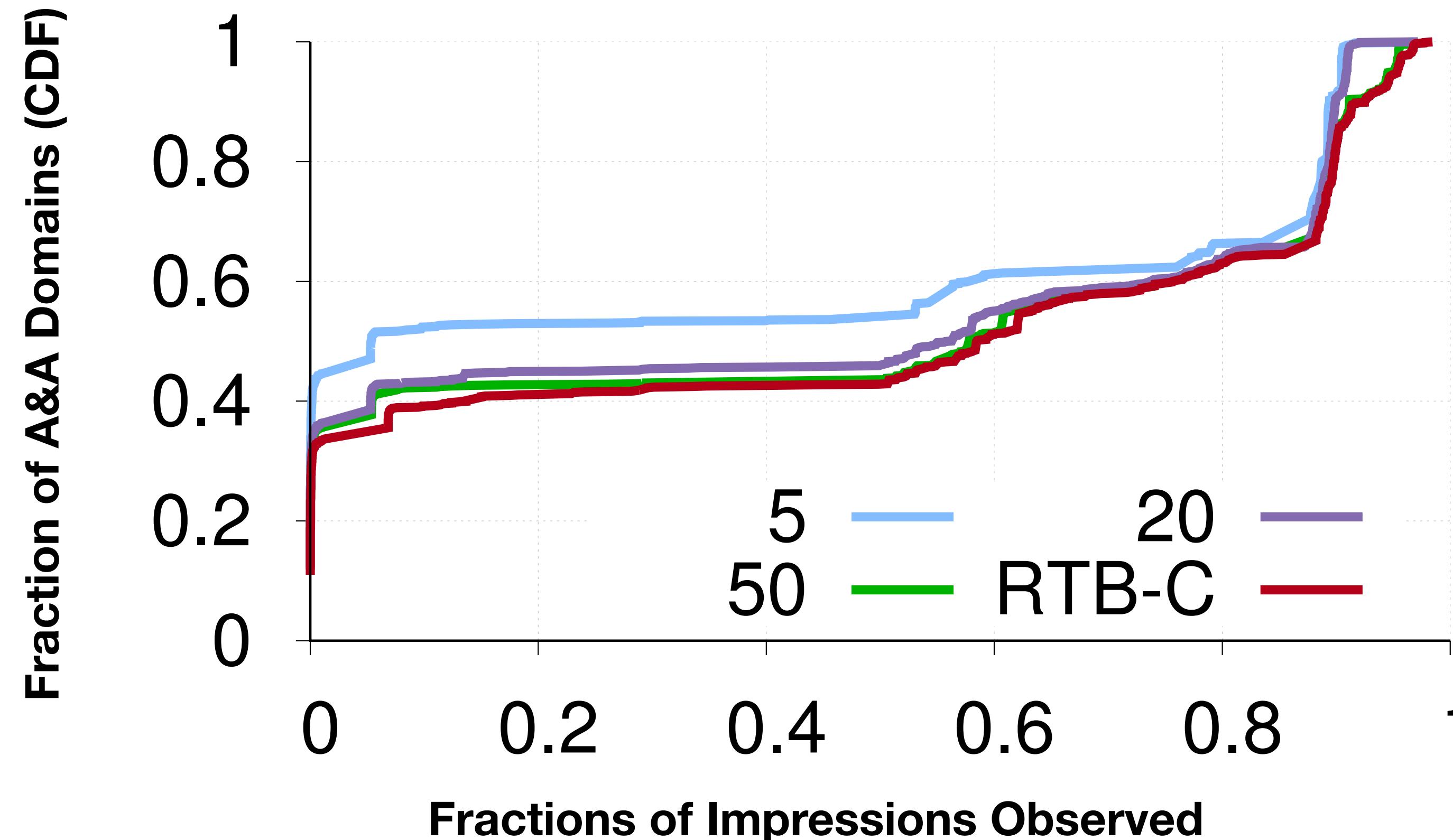


Fig: Fraction of impressions observed by A&A domains under RTB-C model when top x exchanges are selected

Effect of Ad Exchanges



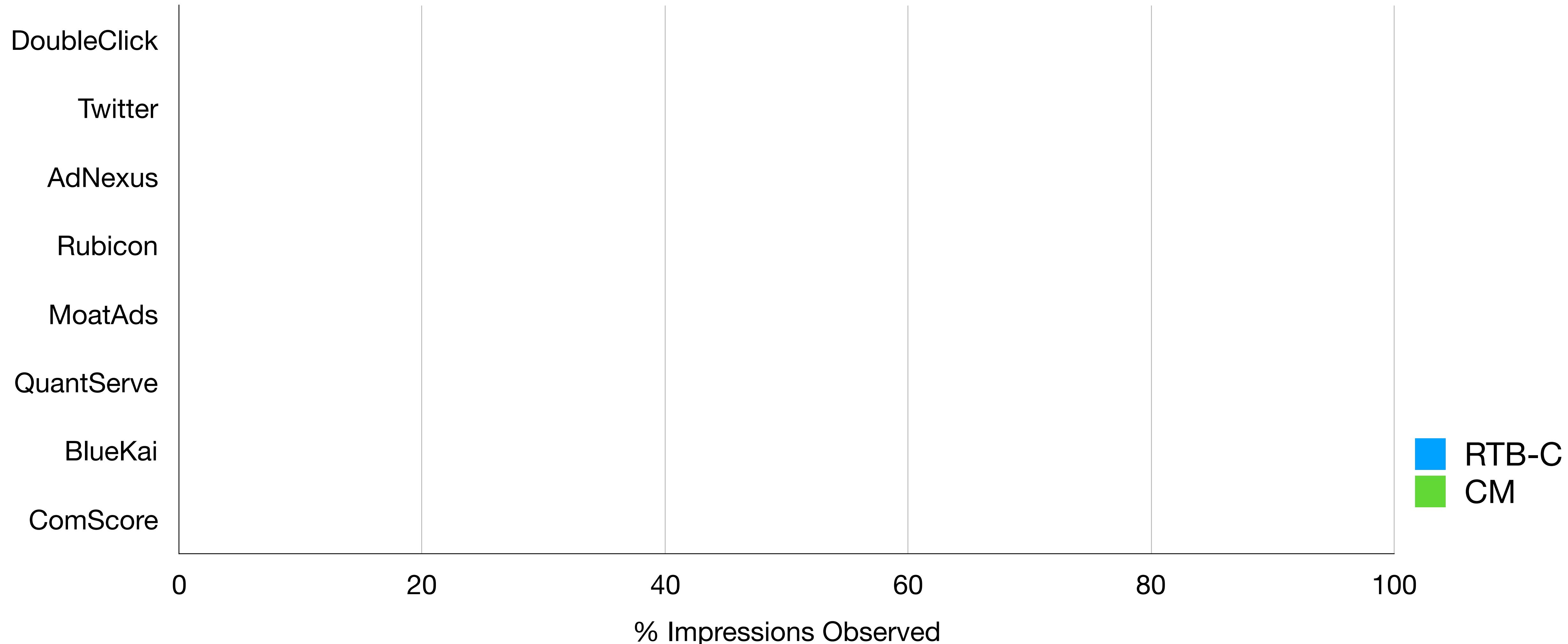
Take Aways

1. RTB-C model is relatively insensitive to the number of exchanges.
2. Graph is extremely dense. Top exchanges are very well-connected.

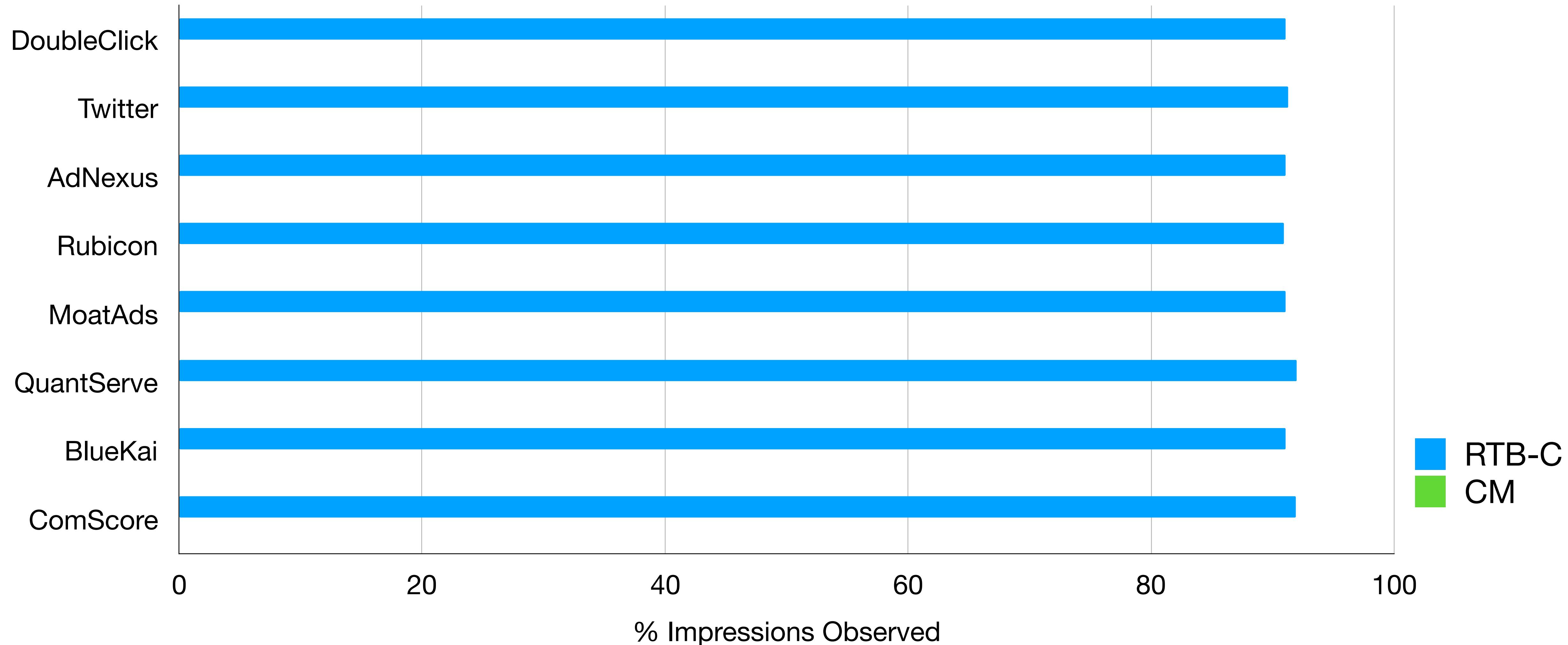
Fig: Fraction of impressions observed by A&A domains under RTB-C model when top x exchanges are selected

Impressions Observed by Popular A&A Domains

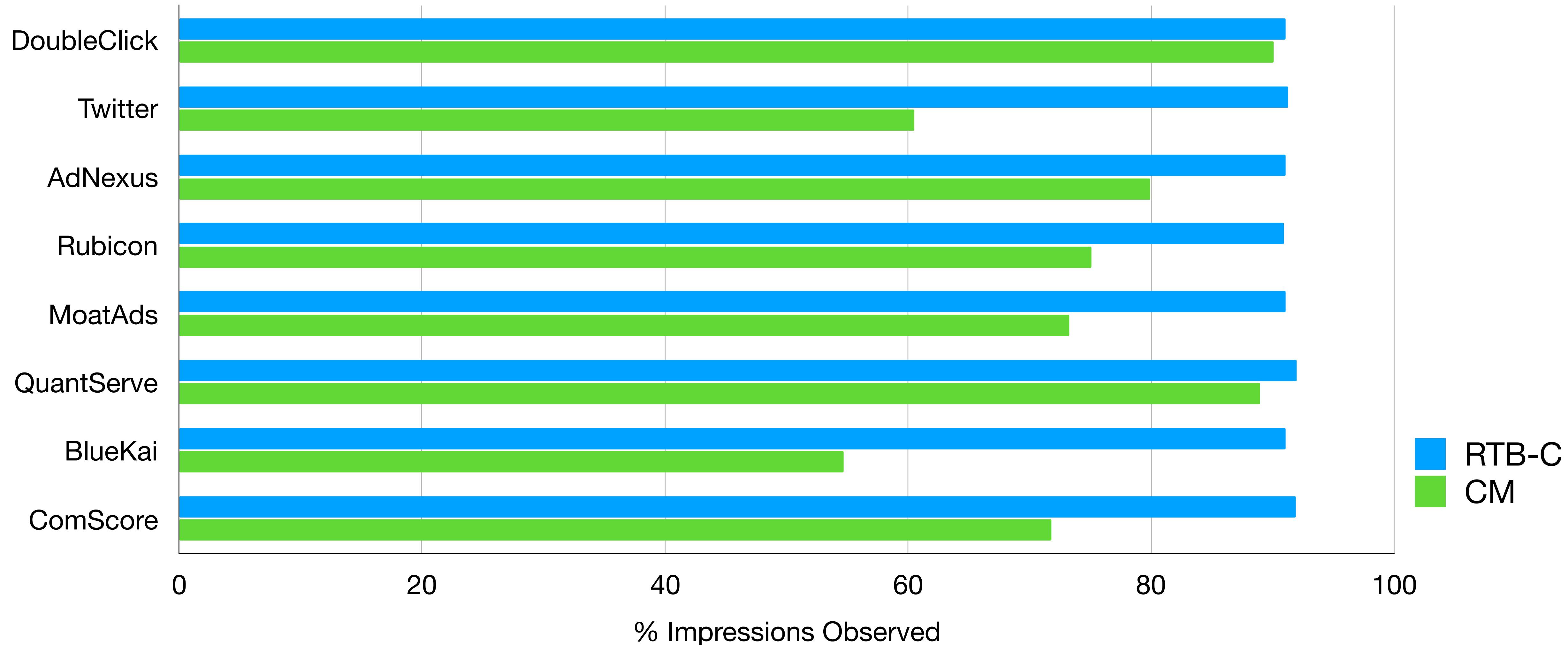
Impressions Observed by Popular A&A Domains



Impressions Observed by Popular A&A Domains



Impressions Observed by Popular A&A Domains



**We can now evaluate the effectiveness
of privacy-preserving tools**

Effect of Blocking Extensions

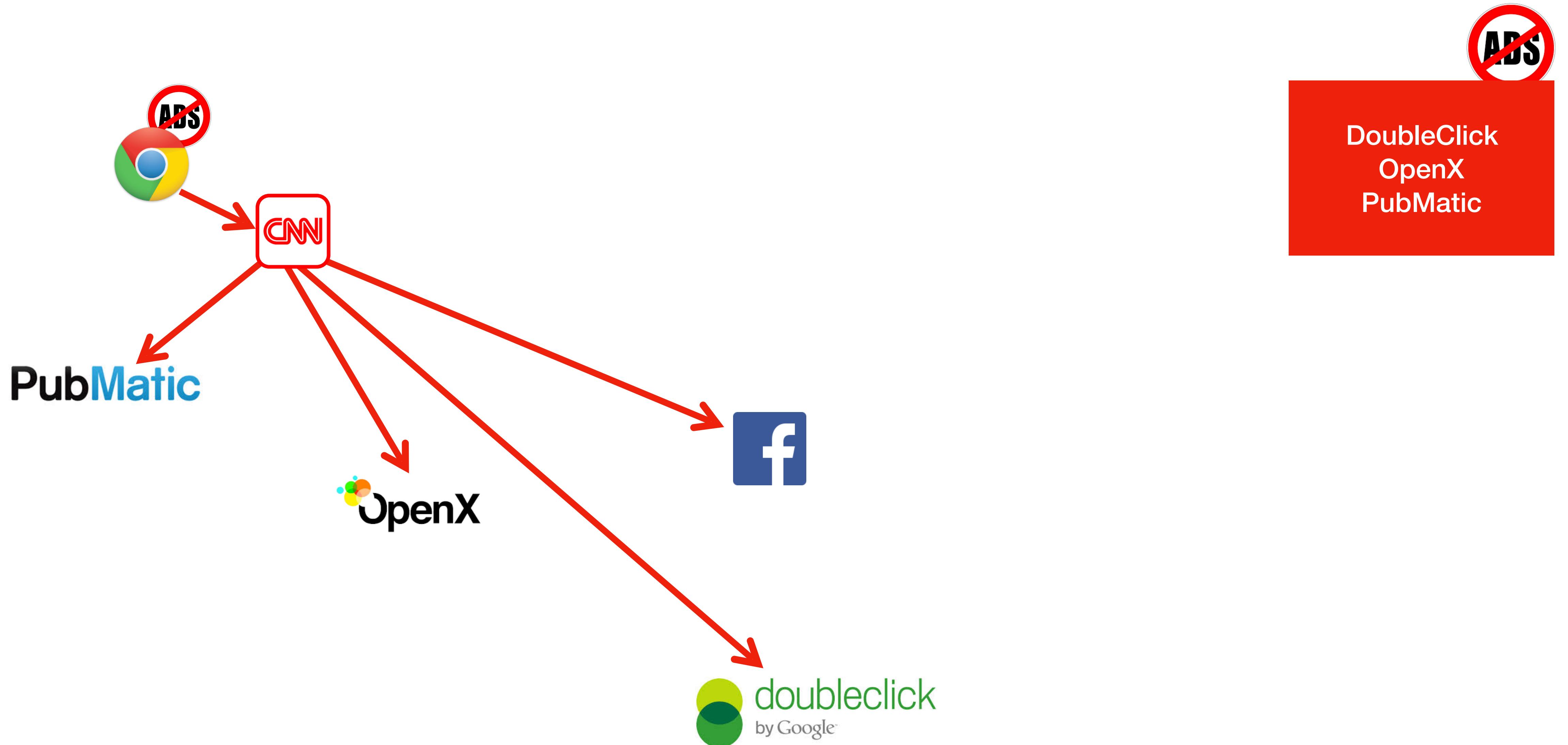


DoubleClick
OpenX
PubMatic

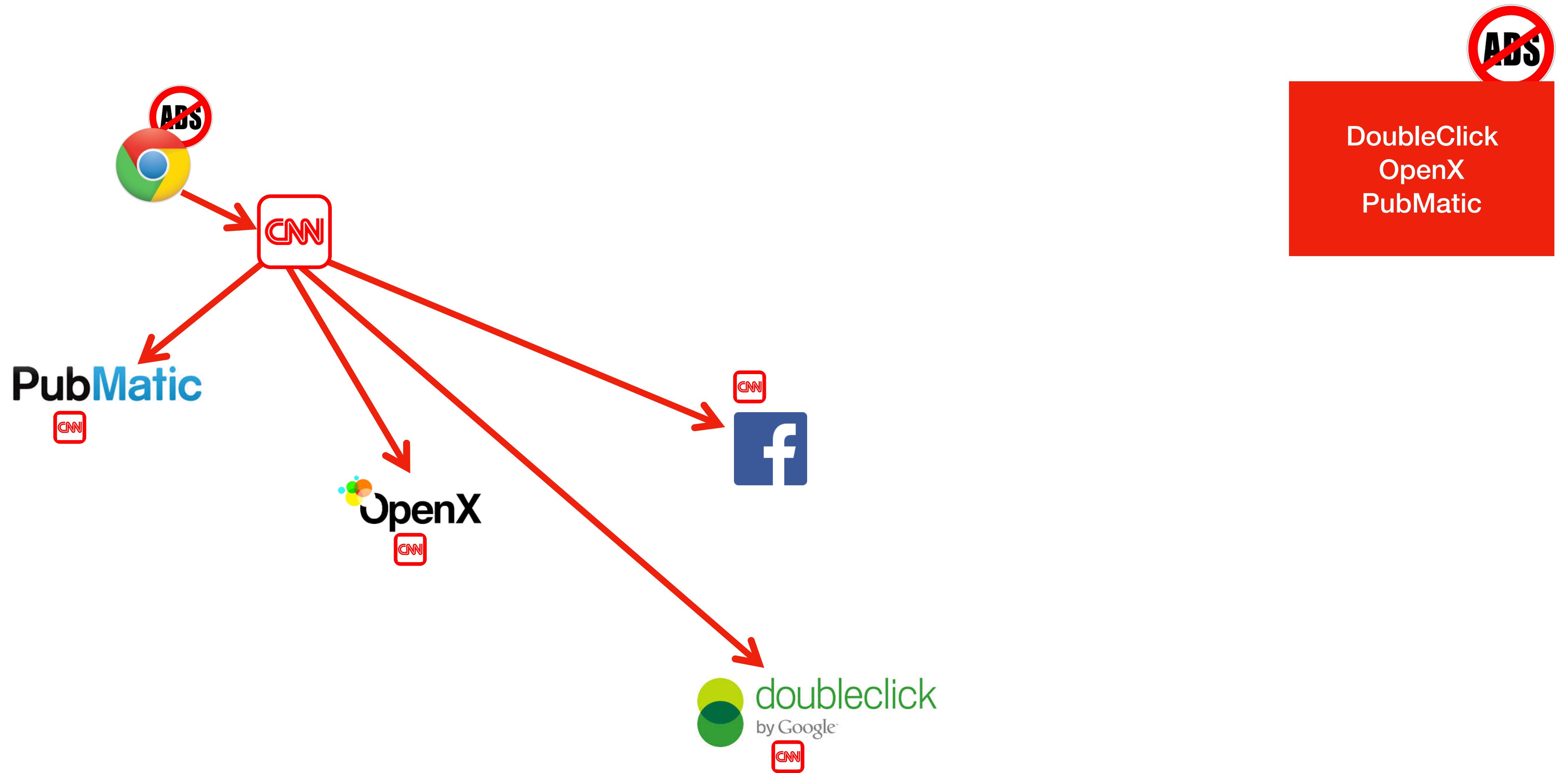
Effect of Blocking Extensions



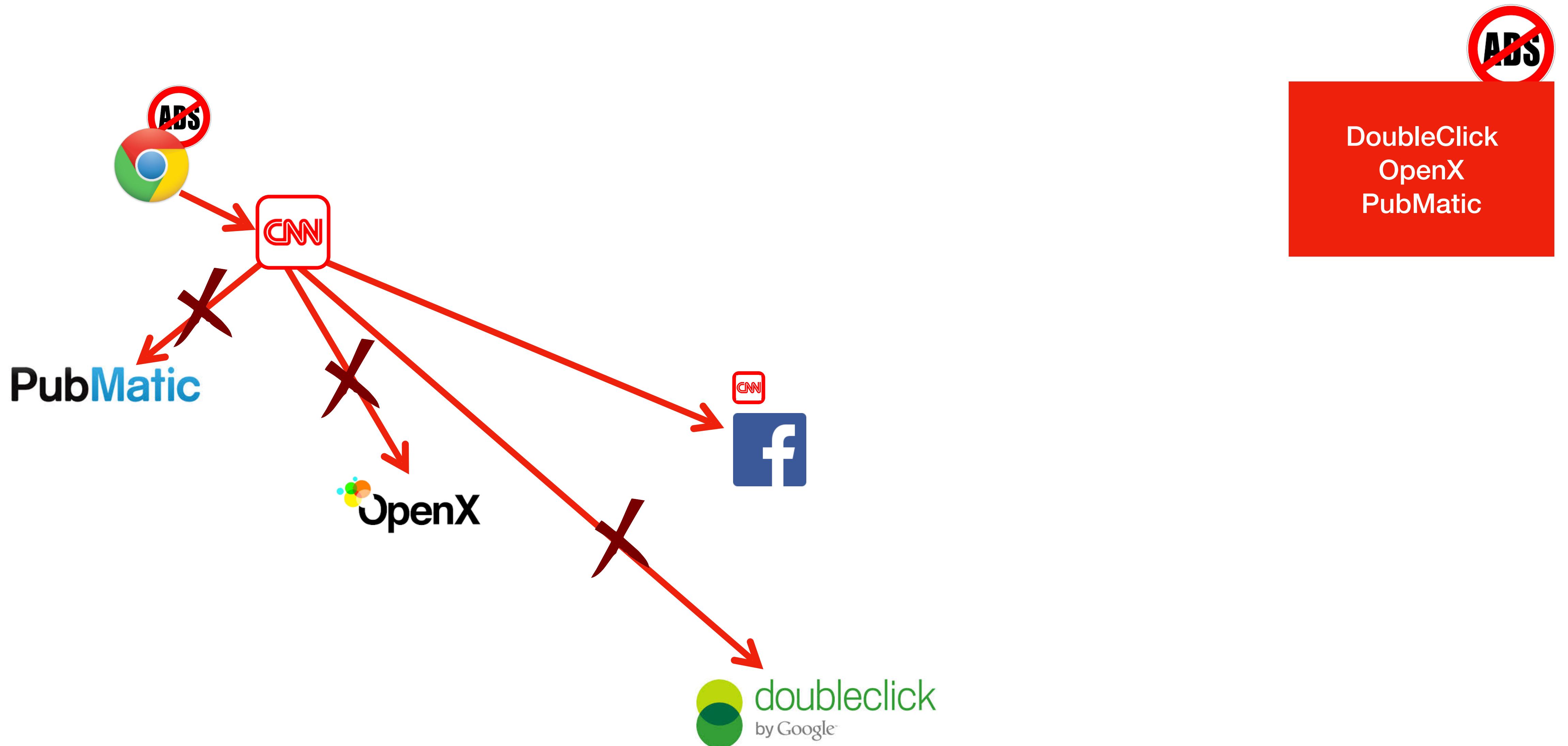
Effect of Blocking Extensions



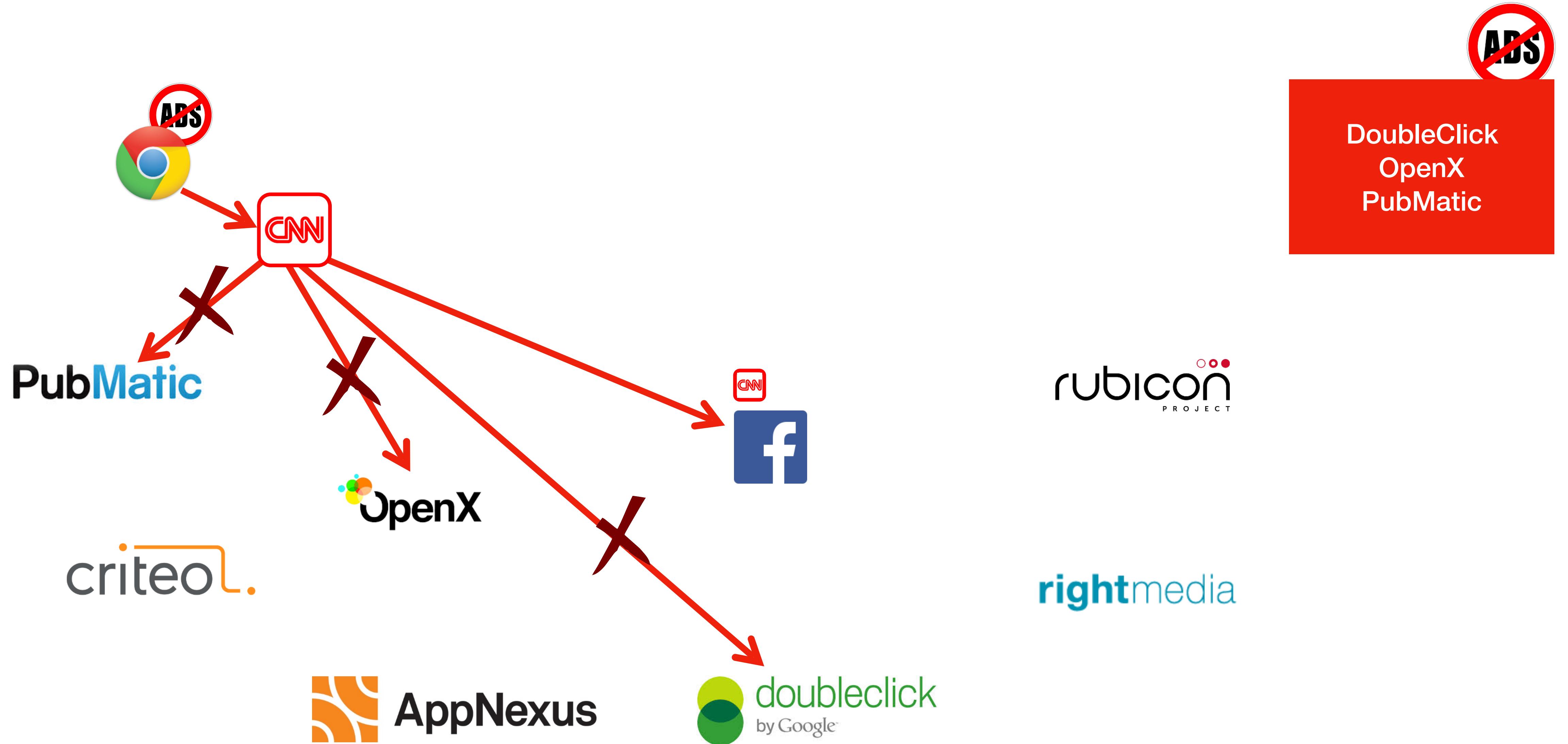
Effect of Blocking Extensions



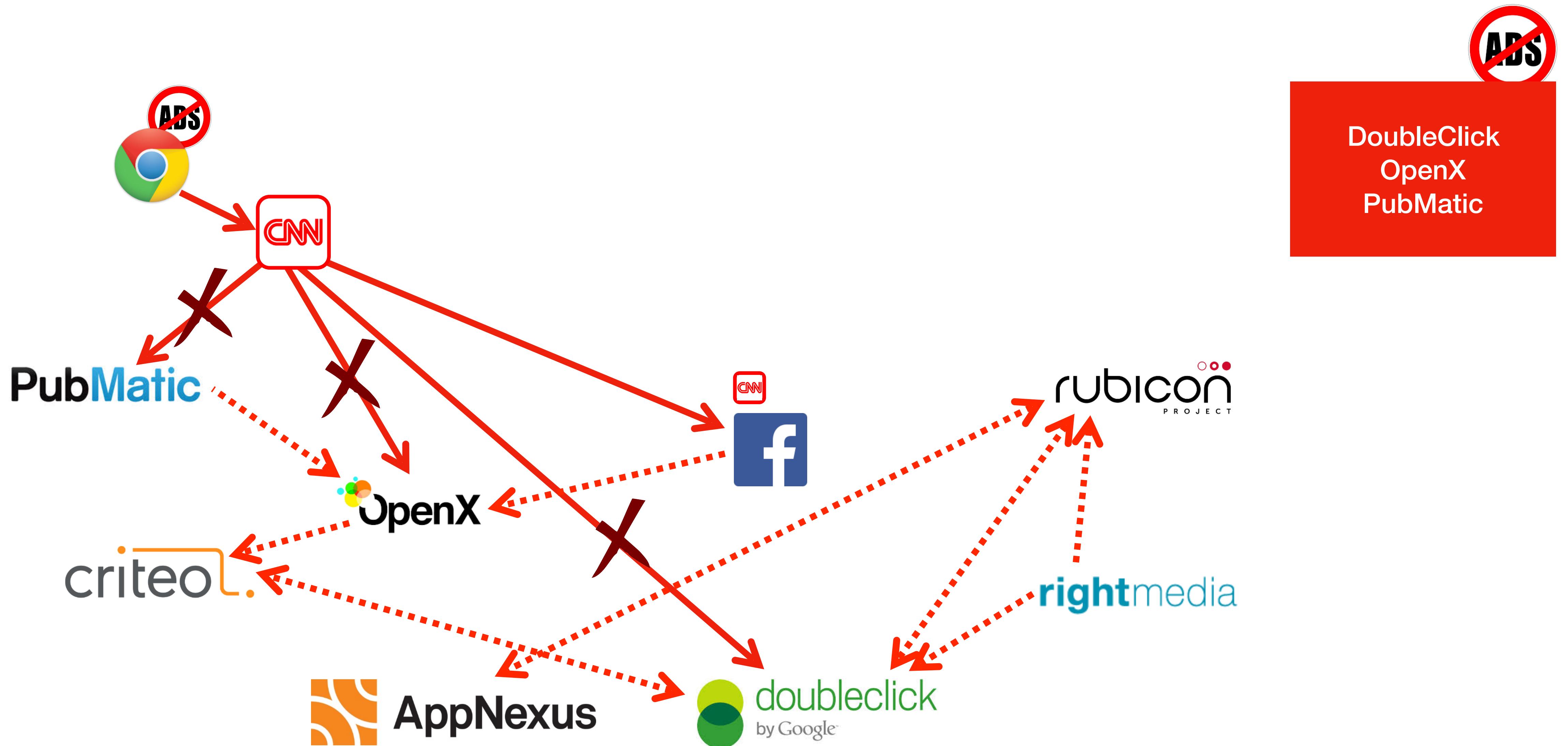
Effect of Blocking Extensions



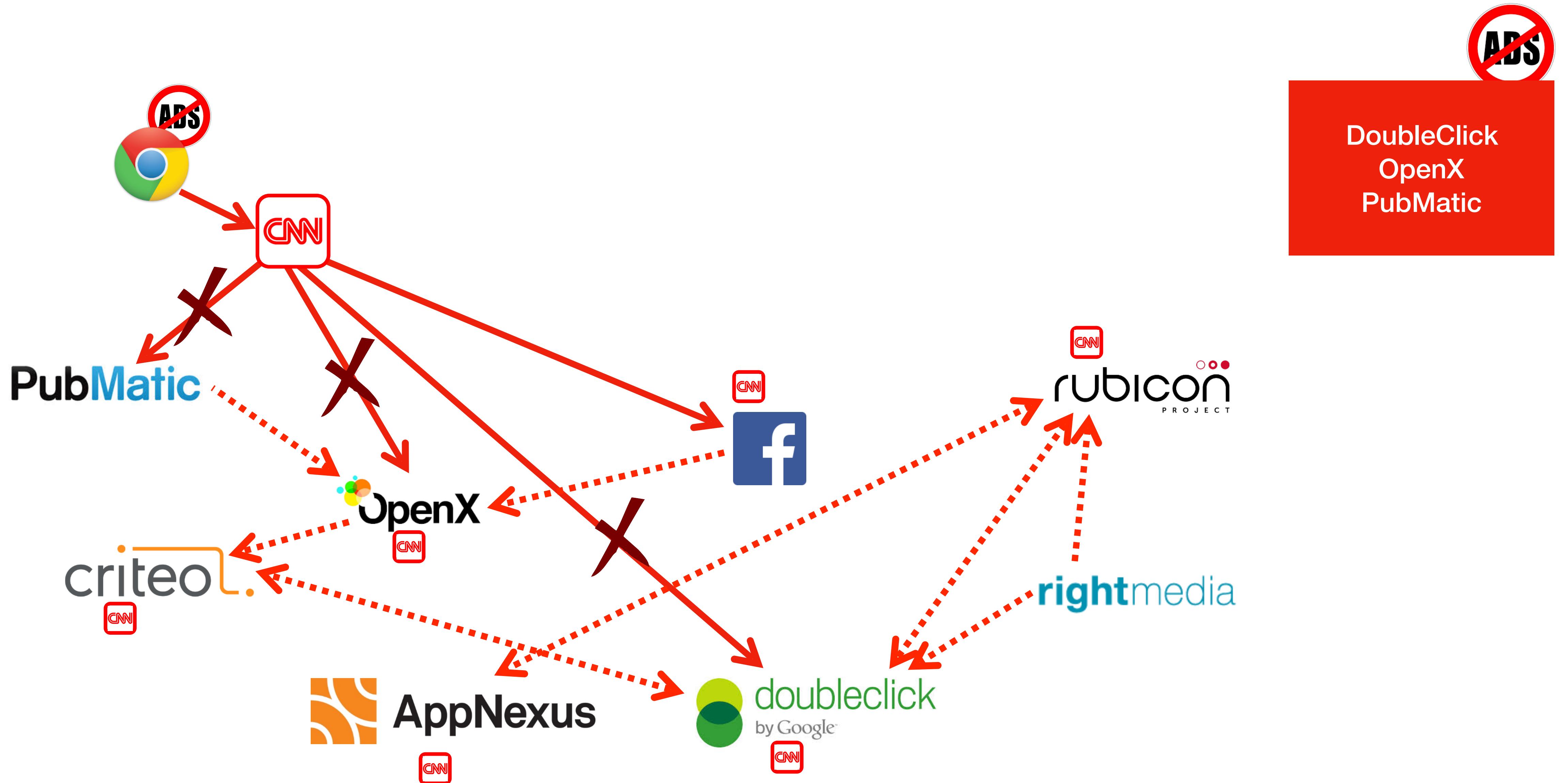
Effect of Blocking Extensions



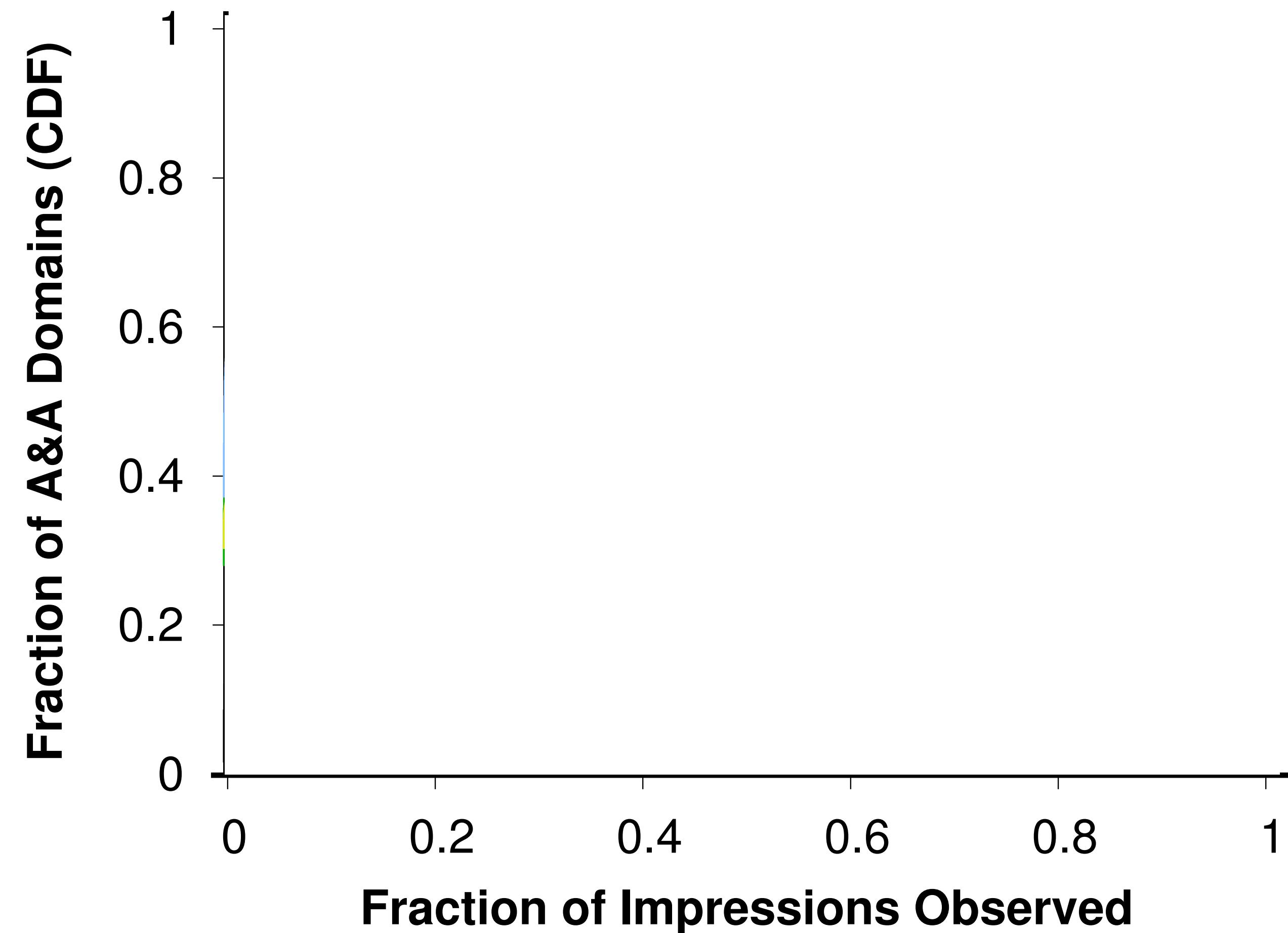
Effect of Blocking Extensions



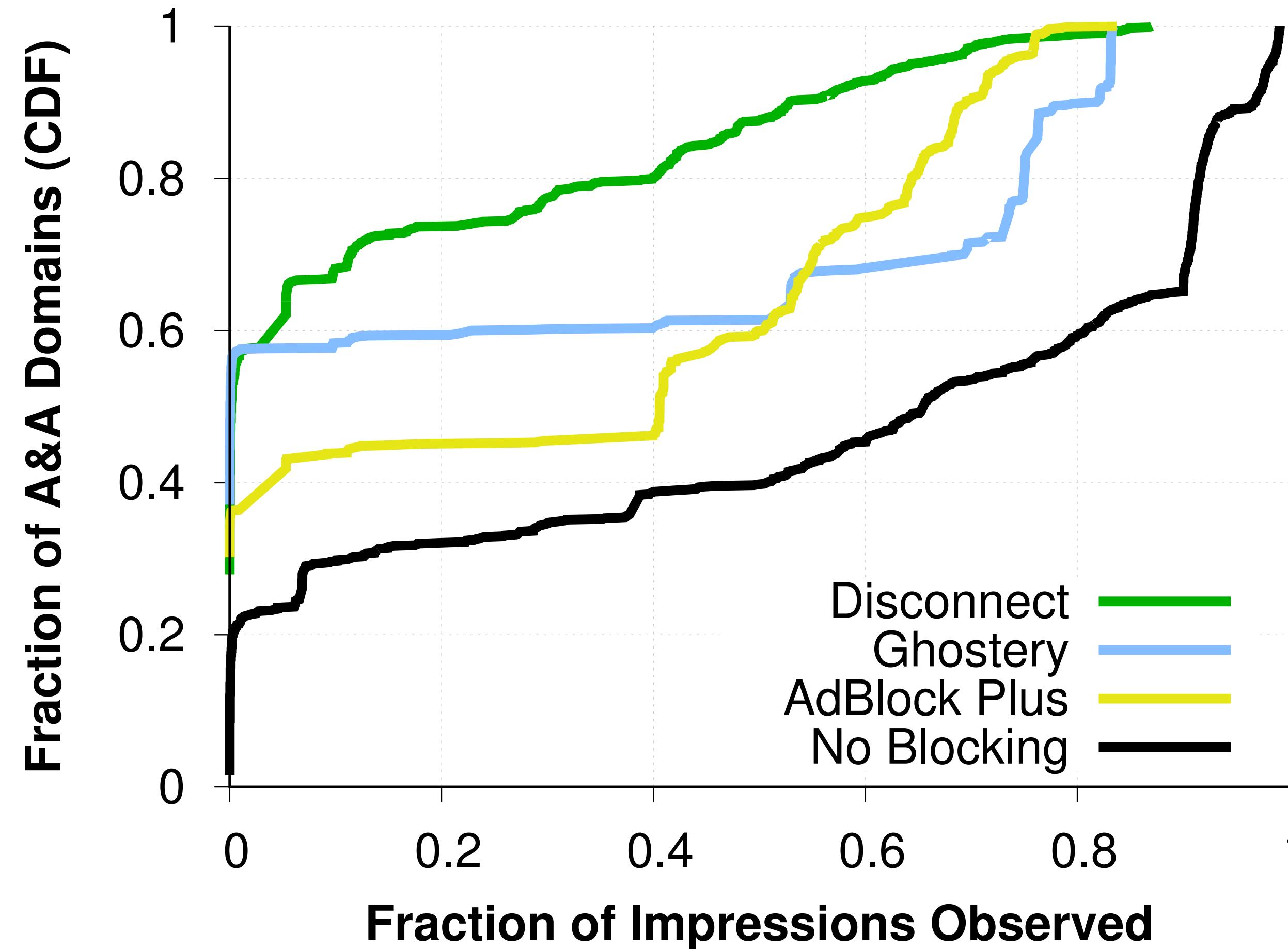
Effect of Blocking Extensions



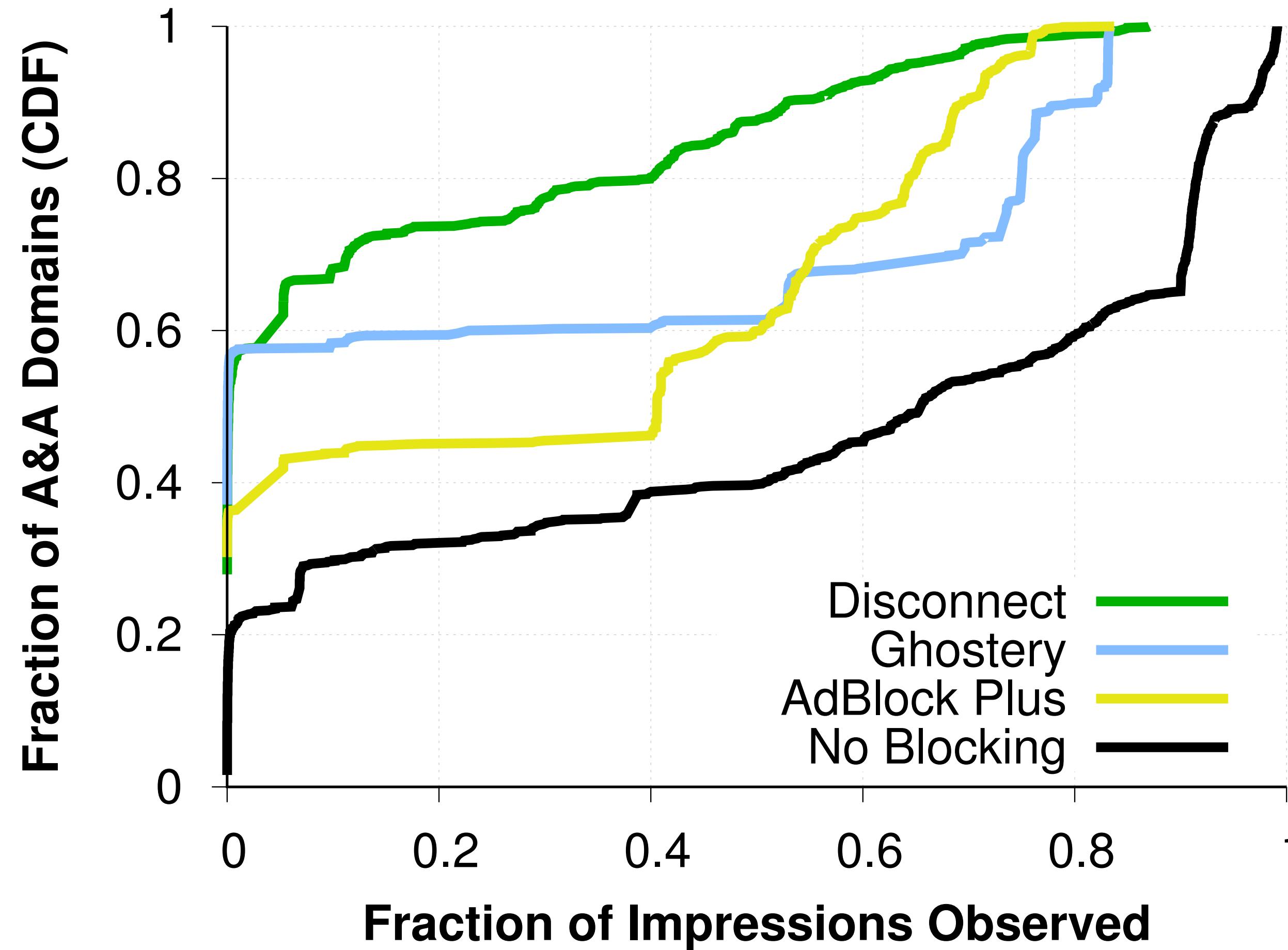
Impressions Observed With Blocking



Impressions Observed With Blocking



Impressions Observed With Blocking

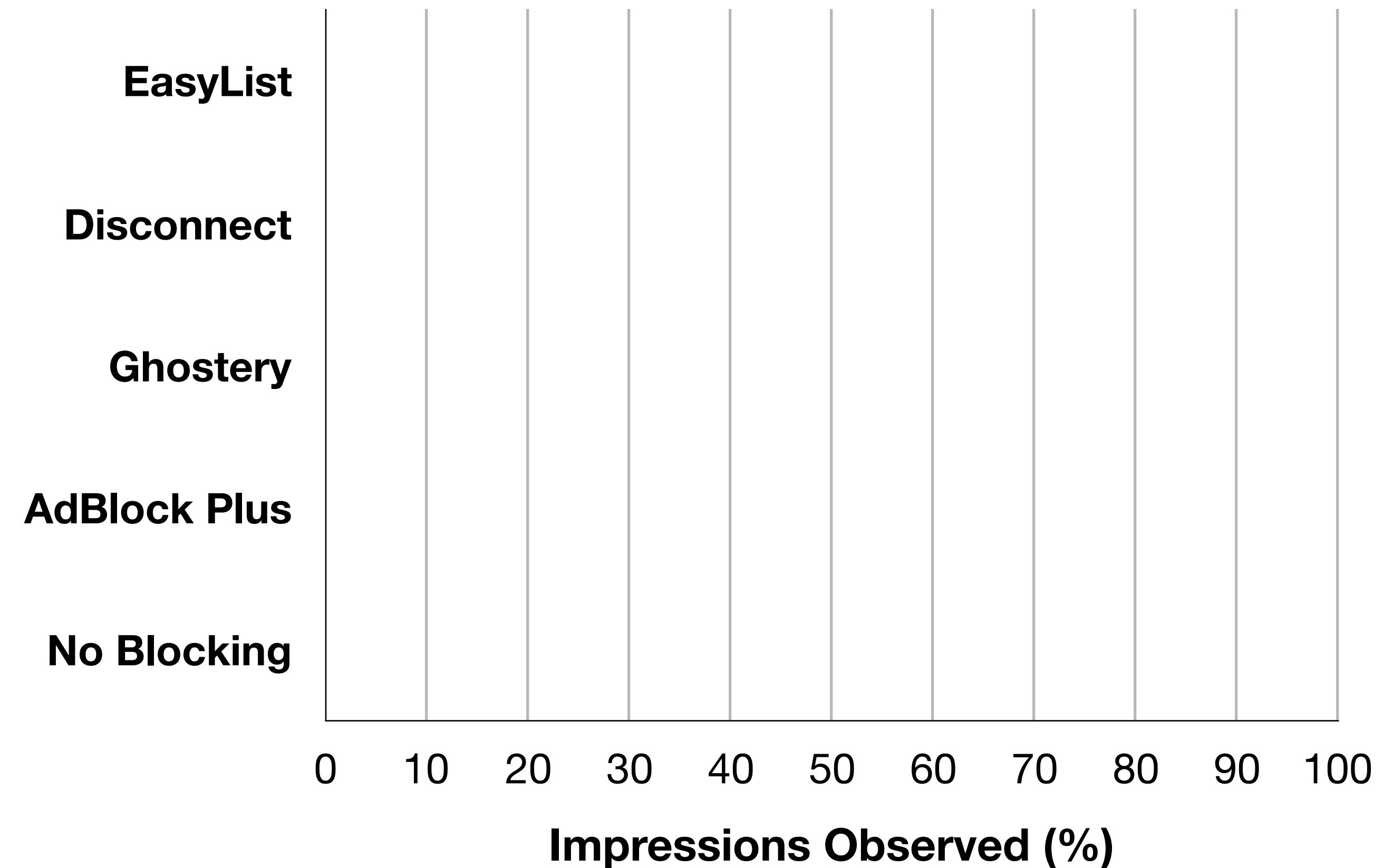


Take Aways

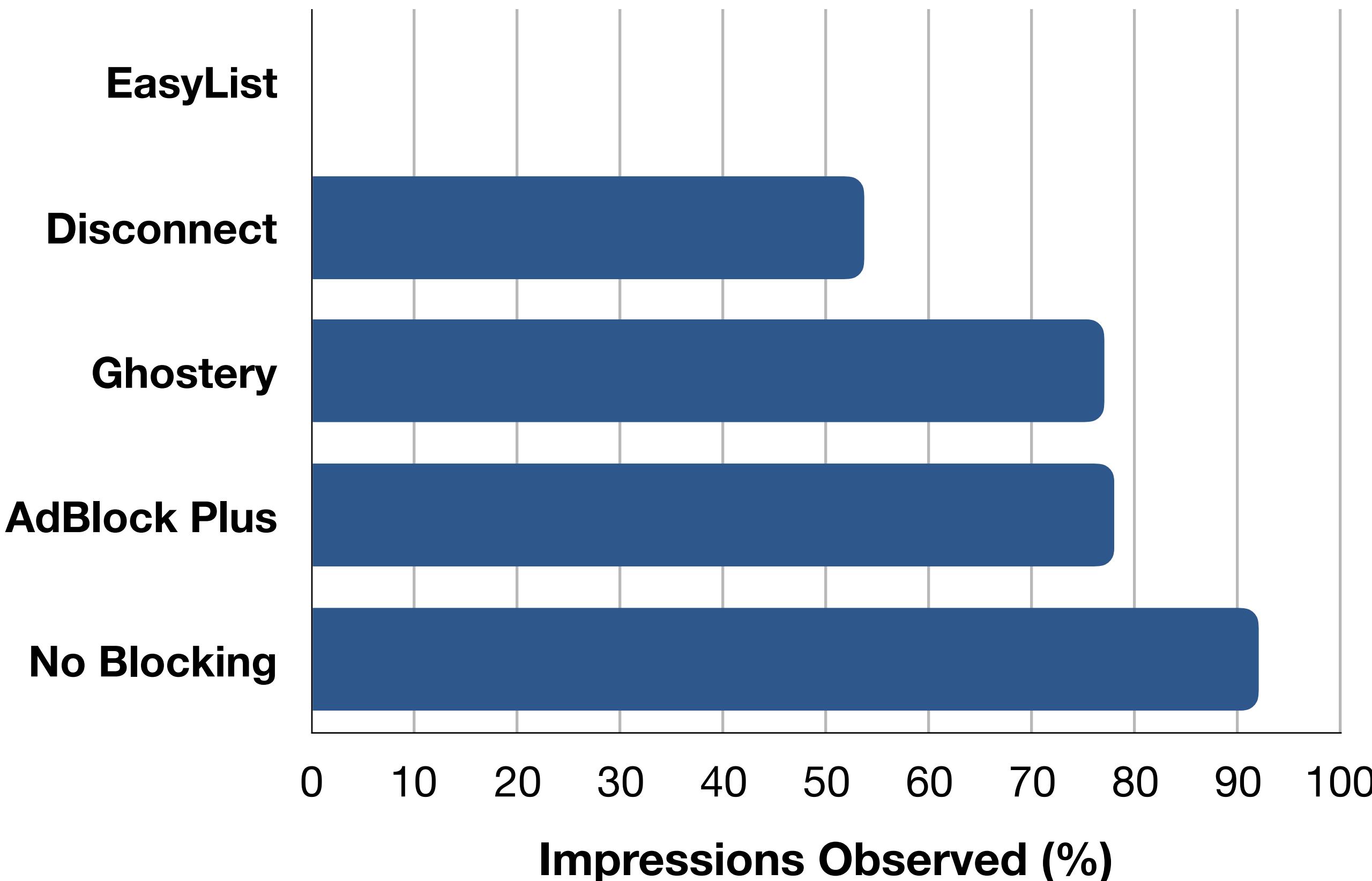
- Disconnect list is most effective.
- ABP is not quite effective due to Acceptable Ads program.
- Due to RTB, impressions are leaked to A&A domains even with blocking extensions.

Top 10 Domains Observing Impressions

Top 10 Domains Observing Impressions



Top 10 Domains Observing Impressions



Top 10 domains can view majority of user impressions even with (most) blocking extensions installed

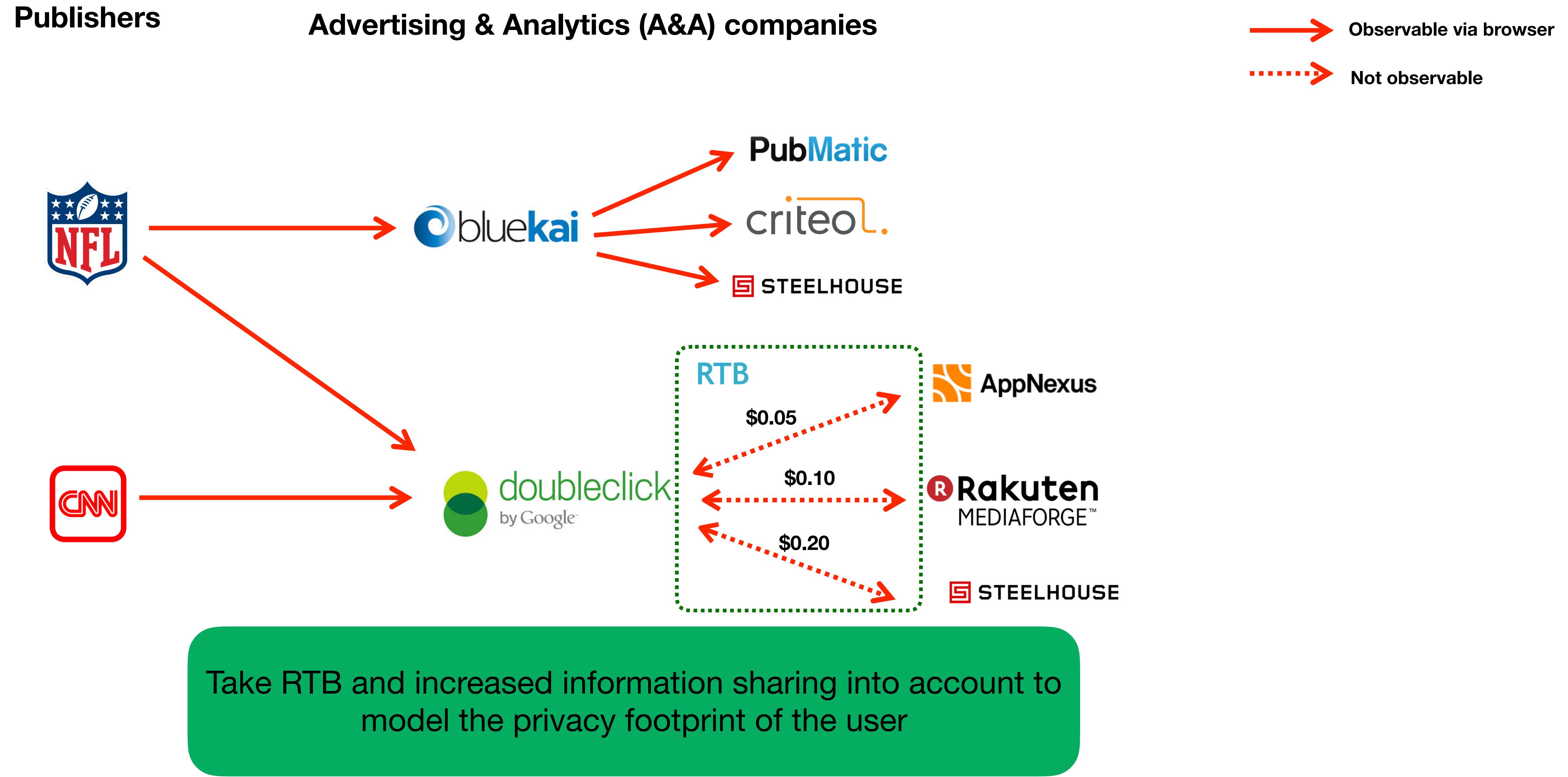
Thesis Outline

Detecting Information Sharing

Isolating List of Ad Exchanges

Modeling User's Privacy Digital Footprint

Have We Modeled the Privacy Footprint Perfectly?



Opportunities to Improve

- Our models provide approximations
 - Need better browsing traces
- The ecosystem could have changed from when the dataset was collected.
 - Longitudinal analysis
 - Geographical patterns
- Not representative of mobile advertising ecosystem

Conclusions

- Provided a content- and platform-agnostic methodology to detect information-sharing between arbitrary A&A domains.
 - Identified cookie-matching partners which would have been missed using prior heuristics.
 - Server-side matching.
 - Raised the bar on how to study information sharing on the web.
- Model to represent the privacy footprint of the user, taking into account RTB.
 - Highlights how information spreads in the advertising graph.
 - Effectiveness of privacy extensions.
 - Even under strict blocking scenarios, top A&A domains can see majority of user impressions.

Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson [Usenix Security 2016]

A First Look at Content Recommendation Networks

Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson [IMC 2016]

Diffusion of User Tracking Data in the Online Advertising Ecosystem

Muhammad Ahmad Bashir and Christo Wilson [PETS 2018]

How Tracking Companies Circumvented Ad Blockers Using WebSockets

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson [IMC 2018]

Evaluating User Interest Profiles Using Ad Preference Managers

Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar and Christo Wilson [NDSS 2019]

A Longitudinal Analysis of the ads.txt Standard

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson. [IMC 2019]

<https://personalization.ccs.neu.edu>

Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson [Usenix Security 2016]

A First Look at Content Recommendation Networks

Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson [IMC 2016]

Diffusion of User Tracking Data in the Online Advertising Ecosystem

Muhammad Ahmad Bashir and Christo Wilson [PETS 2018]

How Tracking Companies Circumvented Ad Blockers Using WebSockets

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson [IMC 2018]

Evaluating User Interest Profiles Using Ad Preference Managers

Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar and Christo Wilson [NDSS 2019]

A Longitudinal Analysis of the ads.txt Standard

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson. [IMC 2019]

<https://personalization.ccs.neu.edu>

Impact

Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson [Usenix Security 2016]

A First Look at Content Recommendation Networks

Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson [IMC 2016]

Diffusion of User Tracking Data in the Online Advertising Ecosystem

Muhammad Ahmad Bashir and Christo Wilson [PETS 2018]

How Tracking Companies Circumvented Ad Blockers Using WebSockets

Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson [IMC 2018]

Evaluating User Interest Profiles Using Ad Preference Managers

Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar and Christo Wilson [NDSS 2019]

A Longitudinal Analysis of the ads.txt Standard

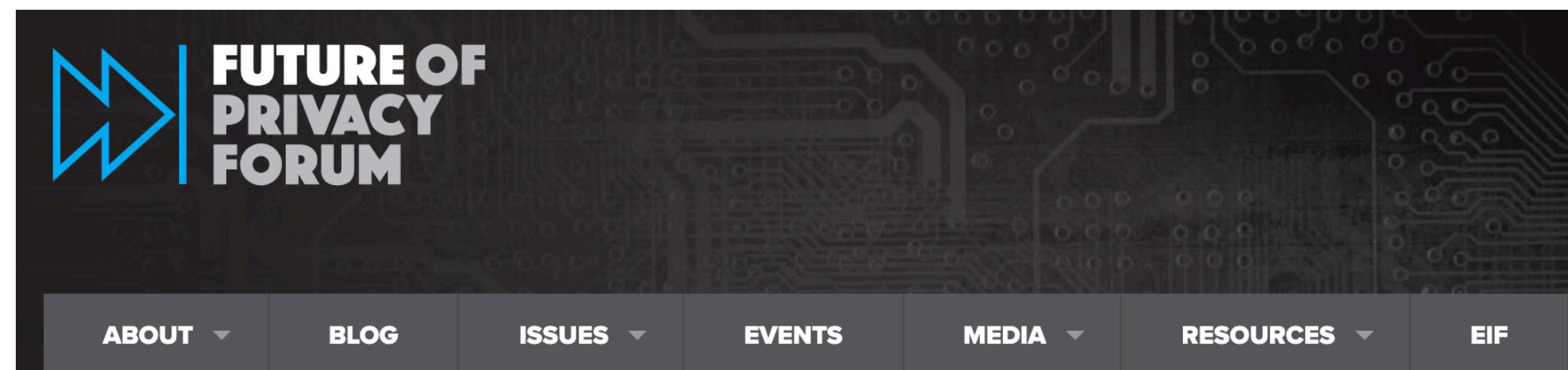
Muhammad Ahmad Bashir, Sajjad Arshad, Engin Kirda, William Robertson, Christo Wilson. [IMC 2019]

<https://personalization.ccs.neu.edu>

FPF Best Student Paper Award

Diffusion of User Tracking Data in the Online Advertising Ecosystem

Muhammad Ahmad Bashir and Christo Wilson [PETS 2018]



This Year's Must-Read Privacy Papers: The Future of Privacy Forum Announces Recipients of Annual Privacy Papers for Policymakers Award



DECEMBER 17, 2018 | EDITOR

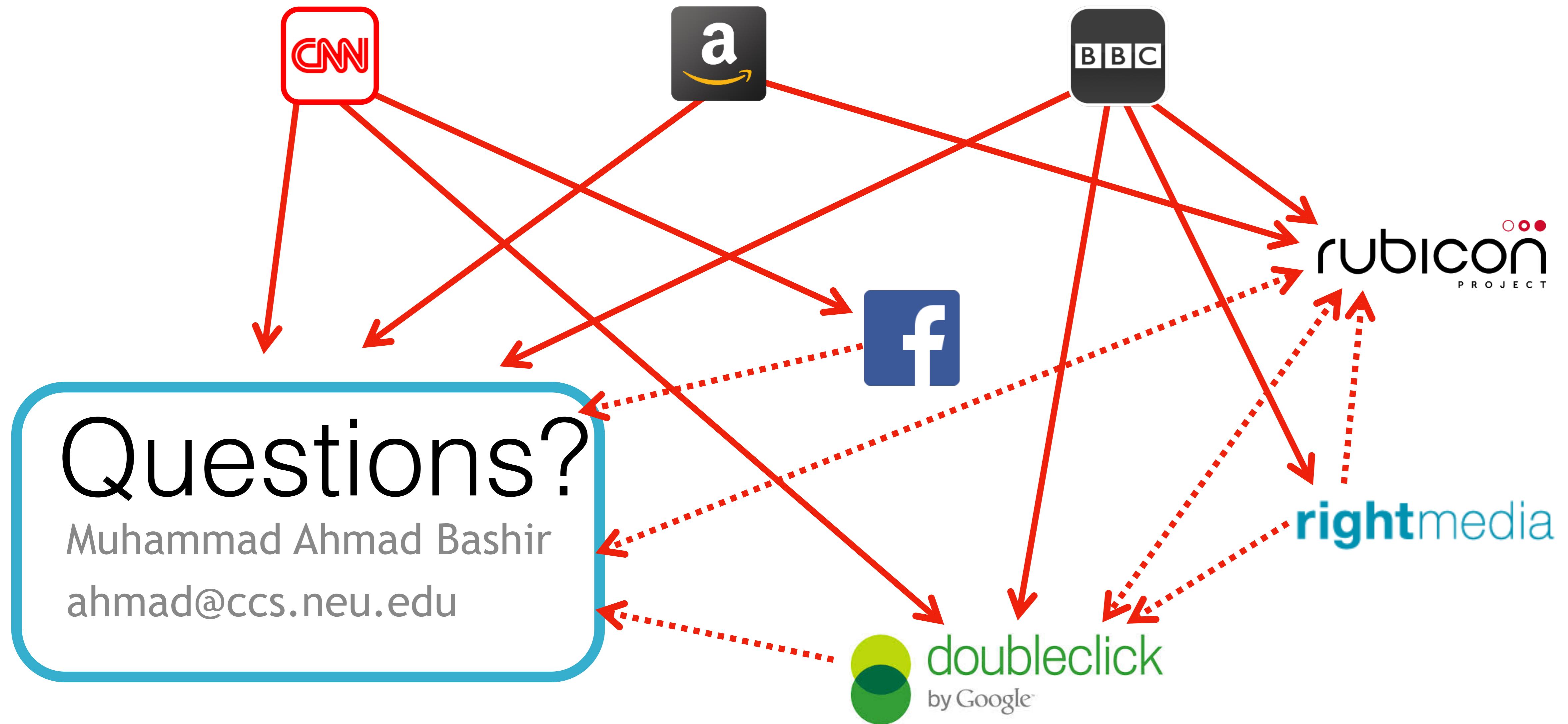


These papers are summarized and distributed to US lawmakers and their staff

Future Directions

- Cross-device tracking
 - Mobile usage increasing + smart devices
- Auditing
 - GDPR compliance
- Other programmatic advertising ecosystems
 - Header Bidding
- Better privacy-enhancing tools

Thank You!



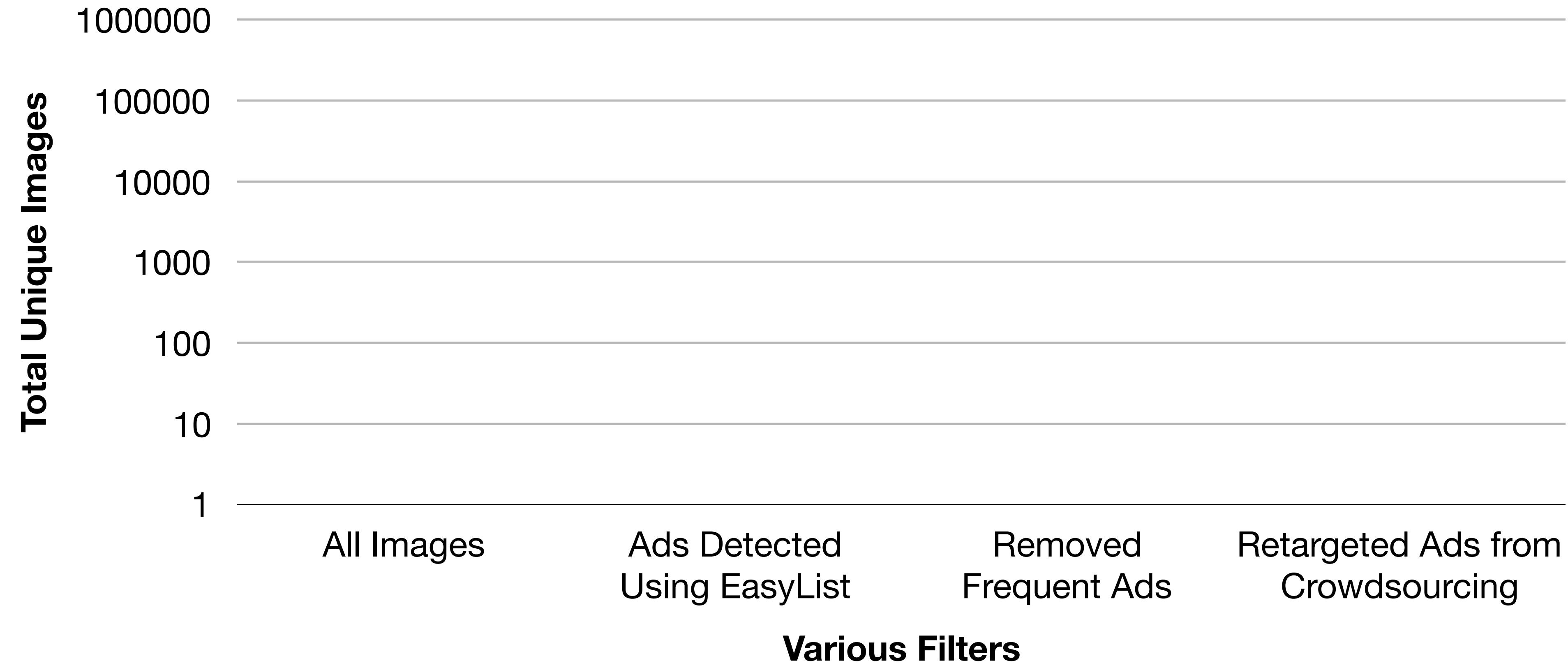
Back up Slides

References

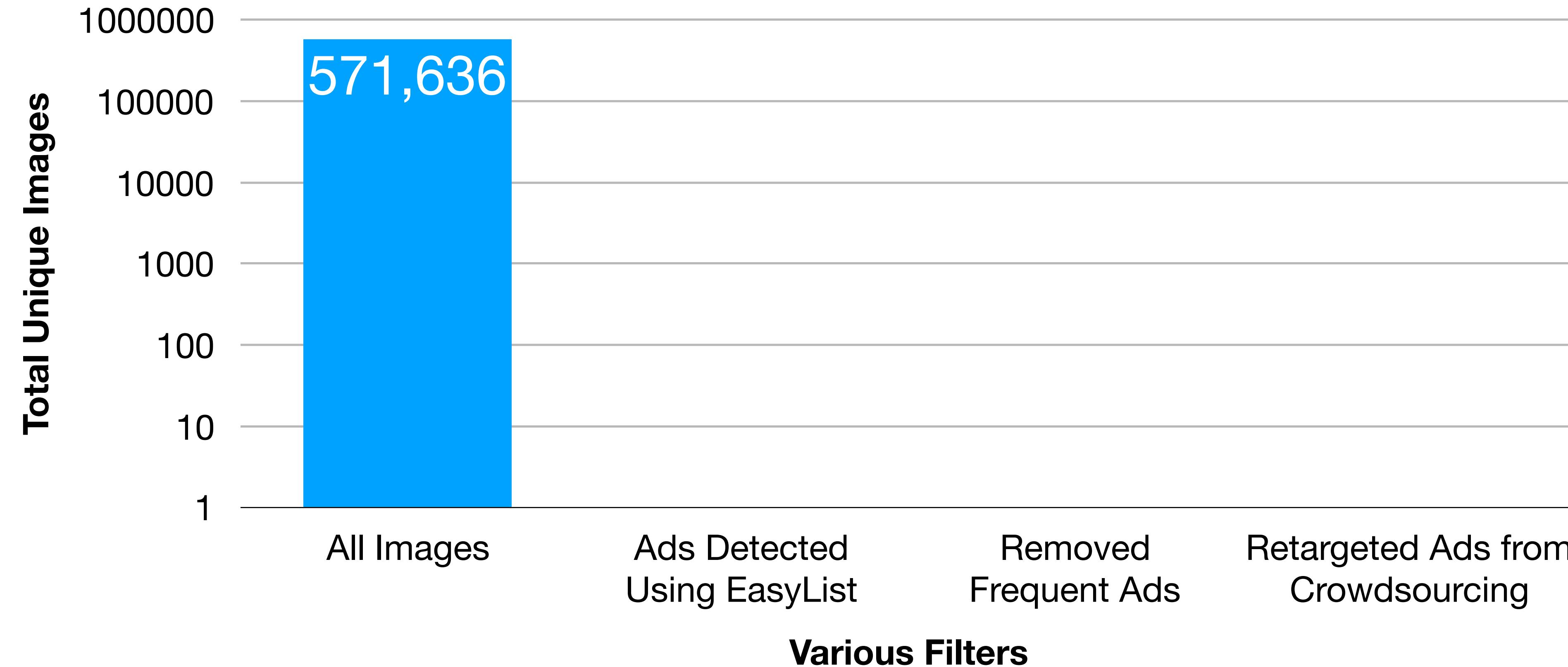
RTB Growth: http://contentb-na2.emarketer.com/content3/US_Programmatic_Ad_Spending_Forecast_2018_eMarketer.pdf

Overall Numbers

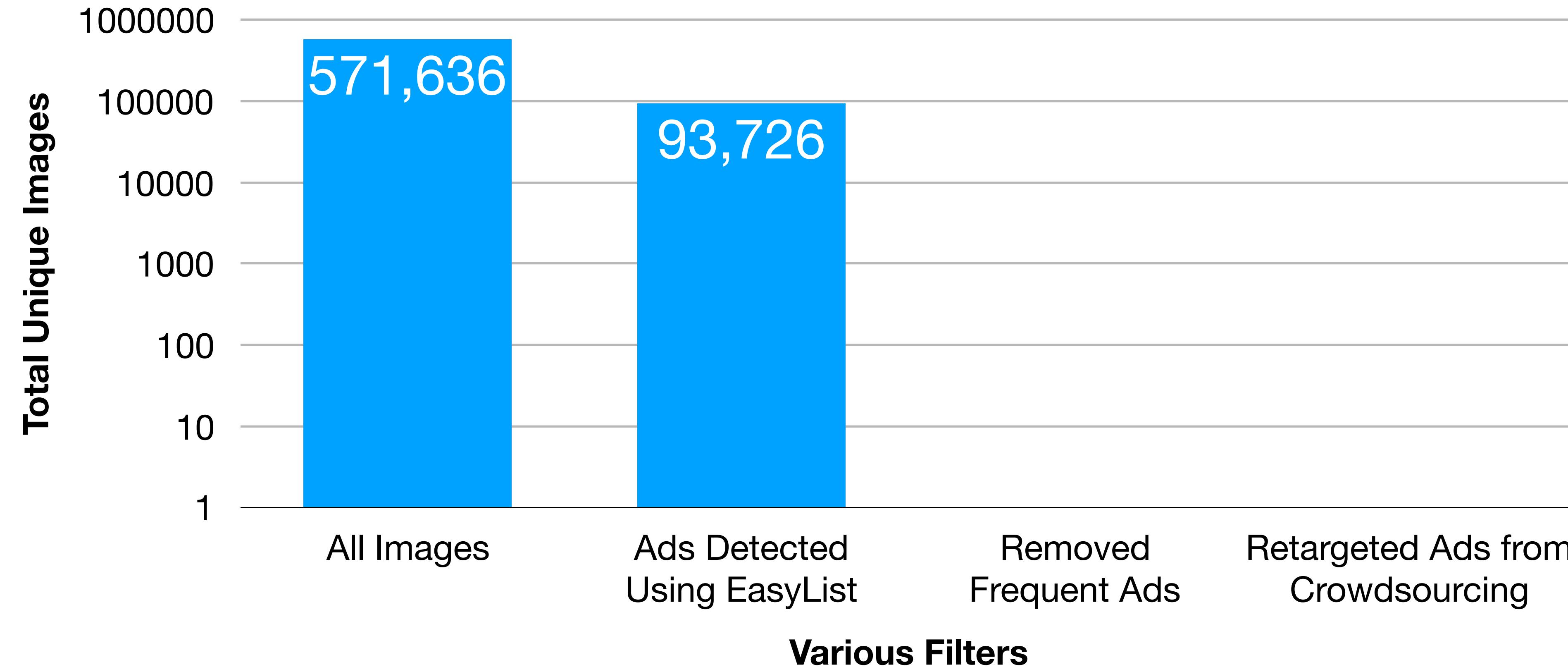
Overall Numbers



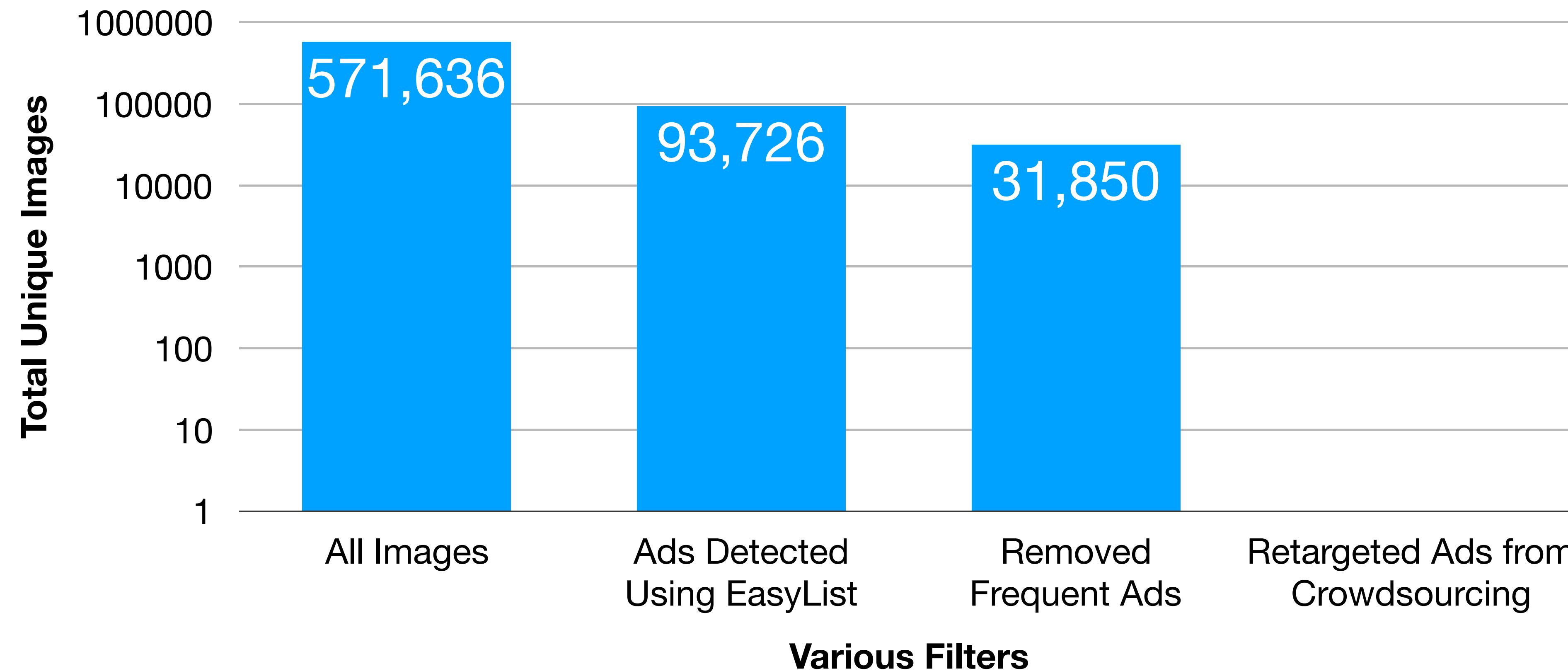
Overall Numbers



Overall Numbers

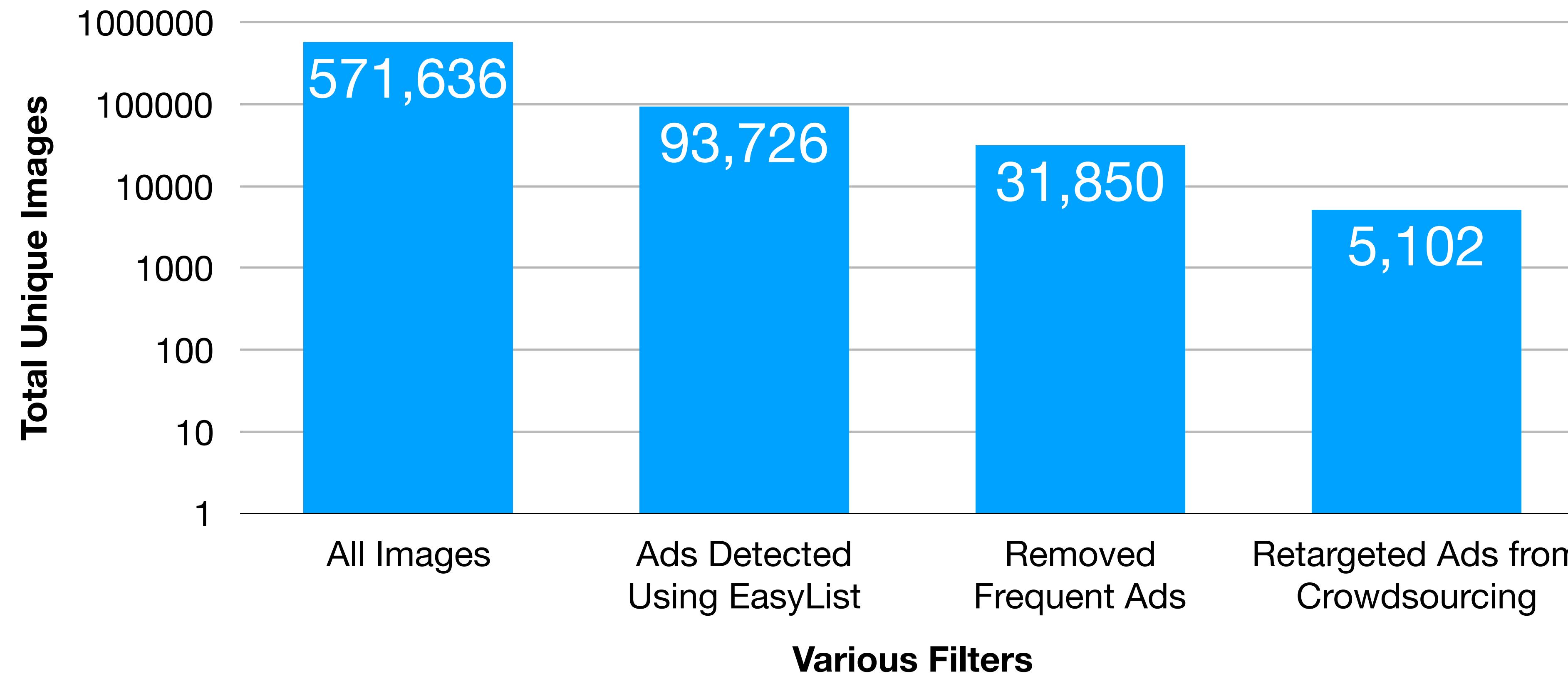


Overall Numbers



- Personas visited non-overlapping retailers
 - By definition, retargets should only be shown to a single persona
- Divided ~32K Images among 1,142 Tasks on Amazon Mechanical Turk (AMT)
 - Each task asked the worker to label 30 ad images

Overall Numbers



- Personas visited non-overlapping retailers
 - By definition, retargets should only be shown to a single persona
- Divided ~32K Images among 1,142 Tasks on Amazon Mechanical Turk (AMT)
 - Each task asked the worker to label 30 ad images

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input checked="" type="radio"/> jewelry <input checked="" type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p>	<p>The image shows an advertisement for Adiamor Diamonds & Fine Jewelry. It features two diamond rings: one with a large central diamond and a smaller diamond on the side, and another with a large central diamond surrounded by smaller diamonds. The text "You Design It • We Build It" is at the top, followed by "Free Shipping • 30-Day Returns". The Adiamor logo is at the bottom left, and a "Shop Now" button is on the right.</p>
<p>Save and Continue</p>	

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p>	<p>The image shows an advertisement for Adiamor Diamonds & Fine Jewelry. It features two diamond rings: one with a large central diamond and a smaller diamond band, and another with a large central diamond surrounded by smaller diamonds. The text "You Design It • We Build It" is at the top, followed by "Free Shipping • 30-Day Returns". The Adiamor logo is at the bottom left, and a "Shop Now" button is on the right.</p>
<p>Save and Continue</p>	

- We know in which persona we got a particular image.

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p> <p>Does this image say it came from one the following websites?</p> <p><input type="radio"/> adiamor.com <input type="radio"/> bluenile.com <input type="radio"/> gillets.com.au <input type="radio"/> lumeradiamonds.com <input type="radio"/> shaneco.com <input type="radio"/> szul.com <input type="radio"/> totaram.com <input type="radio"/> washingtondiamond.com <input type="radio"/> whiteflash.com <input type="radio"/> No</p> <p><input type="button" value="Save and Continue"/></p>	

- We know in which persona we got a particular image.
- We know the shops (websites) we visited per persona.

Example AMT Task

Your response	Image
<p>Select the appropriate category for the image.</p> <p><input checked="" type="radio"/> shopping <input type="radio"/> jewelry <input type="radio"/> diamonds Rings, necklace, etc.</p> <p><input type="radio"/> None of the above</p> <p>Does this image say it came from one the following websites?</p> <p><input checked="" type="radio"/> adiamor.com <input type="radio"/> bluenile.com <input type="radio"/> gillets.com.au <input type="radio"/> lumeradiamonds.com <input type="radio"/> shaneco.com <input type="radio"/> szul.com <input type="radio"/> totaram.com <input type="radio"/> washingtondiamond.com <input type="radio"/> whiteflash.com <input type="radio"/> No</p>	

- We know in which persona we got a particular image.
- We know the shops (websites) we visited per persona.

An Example of Publisher & Shopper Side Chains

An Example of Publisher & Shopper Side Chains

Example

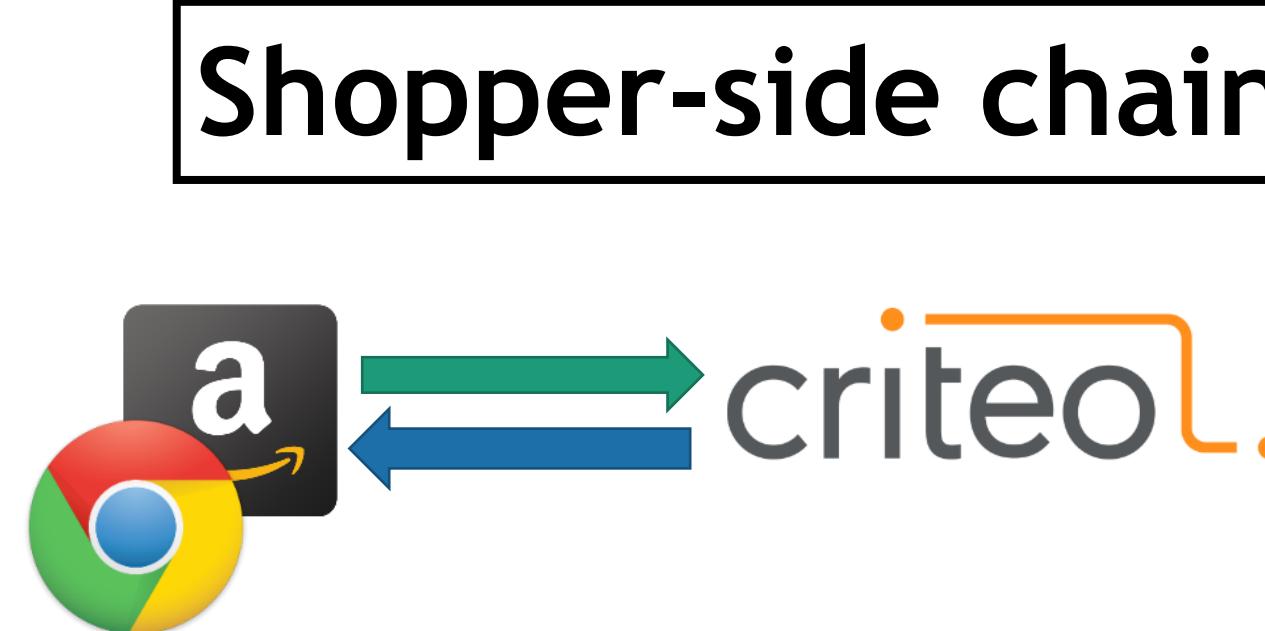
Let's see what caused this Ad to be shown to us?

Publisher-side chain



An Example of Publisher & Shopper Side Chains

Example



Let's see what caused this Ad to be shown to us?

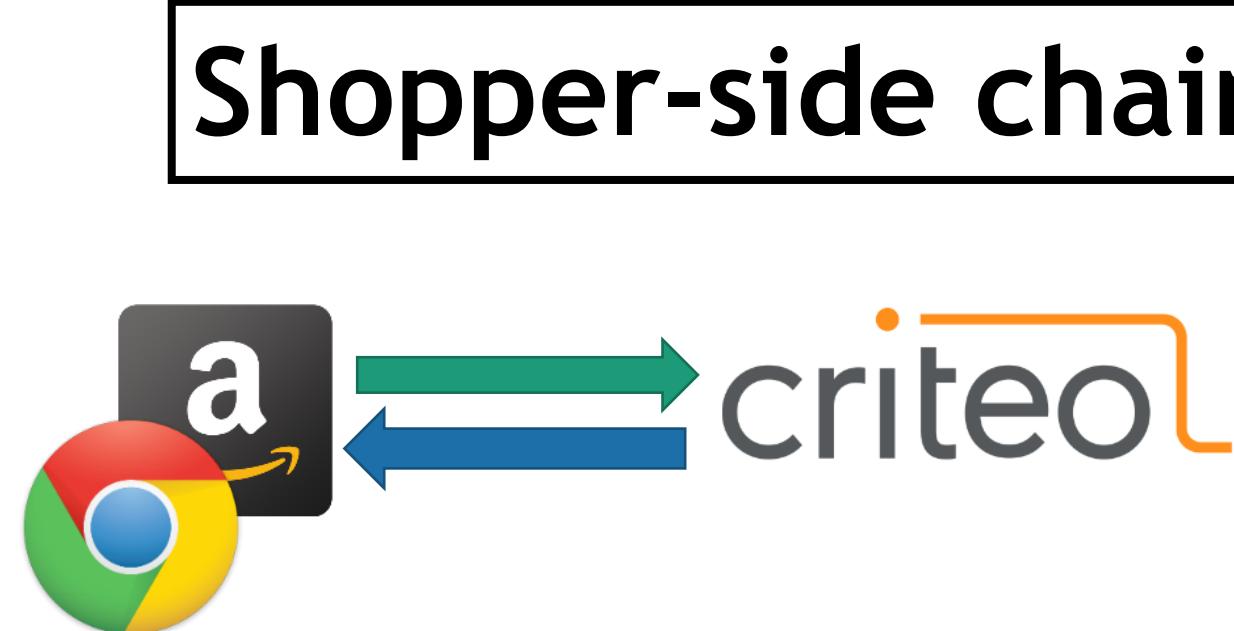


How does criteo know to serve ad on BBC ?

- In this case it is pretty trivial
- criteo observed the user at the (shop) a

An Example of Publisher & Shopper Side Chains

Example



Let's see what caused this Ad to be shown to us?

Publisher-side chain



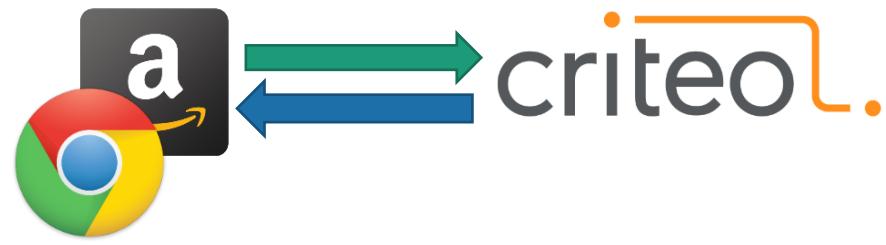
How does criteo know to serve ad on BBC ?

- In this case it is pretty trivial
- criteo observed the user at the (shop) a

Challenge: Can we classify all such publisher-side chains formally?

A Pattern for Chain Classification

Shopper-side chain



Publisher-side chain



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

A Pattern for Chain Classification

Shopper-side chain



Publisher-side chain



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

shop → d

pub → d

A Pattern for Chain Classification

-  d Demand Side Platform (DSP)
-  e Exchange (Running Auction)

Shopper-side chain



Publisher-side chain



shop → d

pub → d

A Pattern for Chain Classification

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Shopper-side chain



Publisher-side chain



shop → d

pub → d

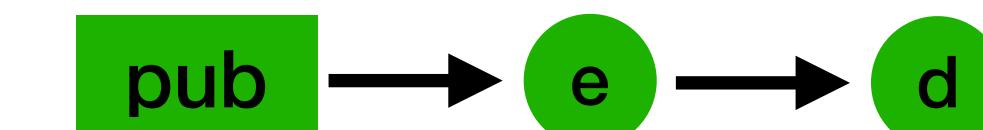
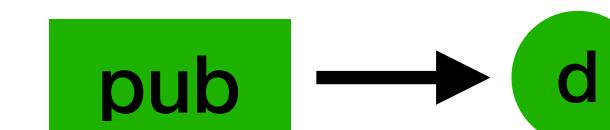
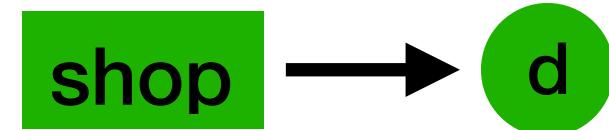
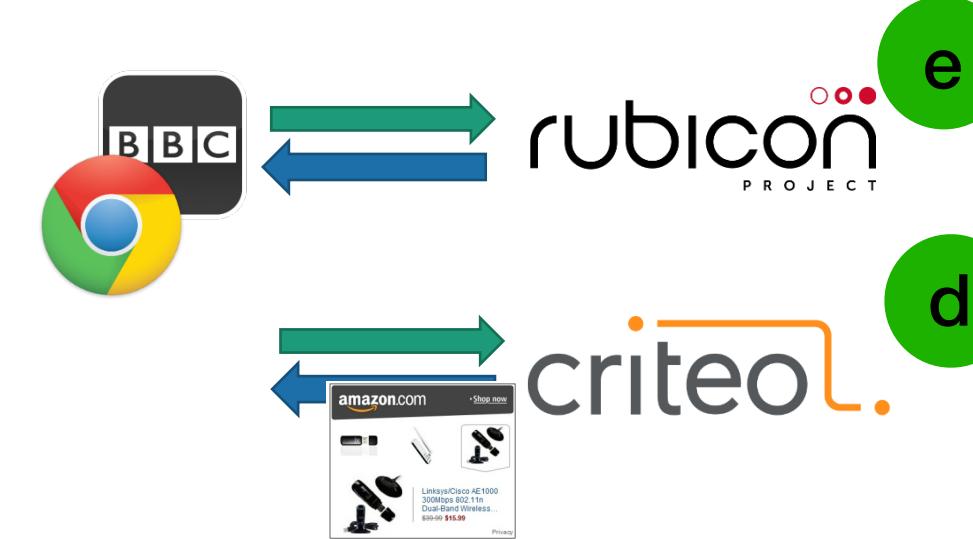
A Pattern for Chain Classification

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Shopper-side chain

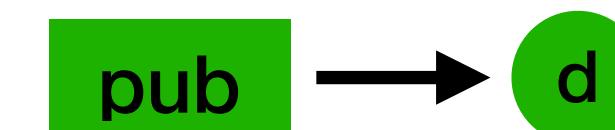
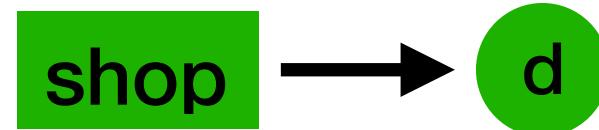


Publisher-side chain



A Pattern for Chain Classification

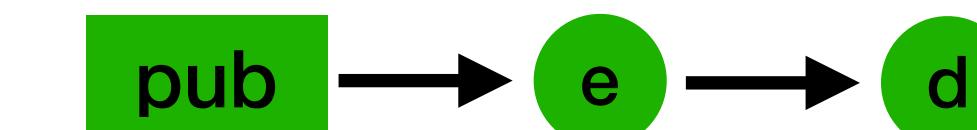
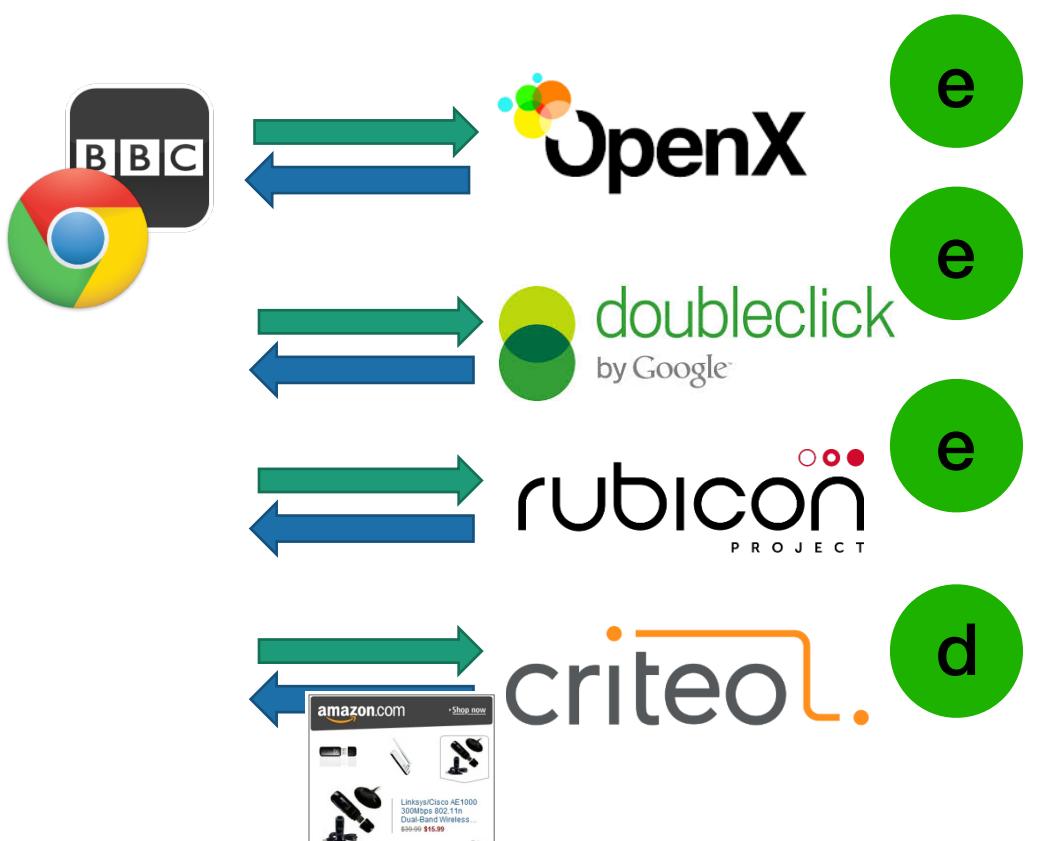
Shopper-side chain



Publisher-side chain



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)



A Pattern for Chain Classification

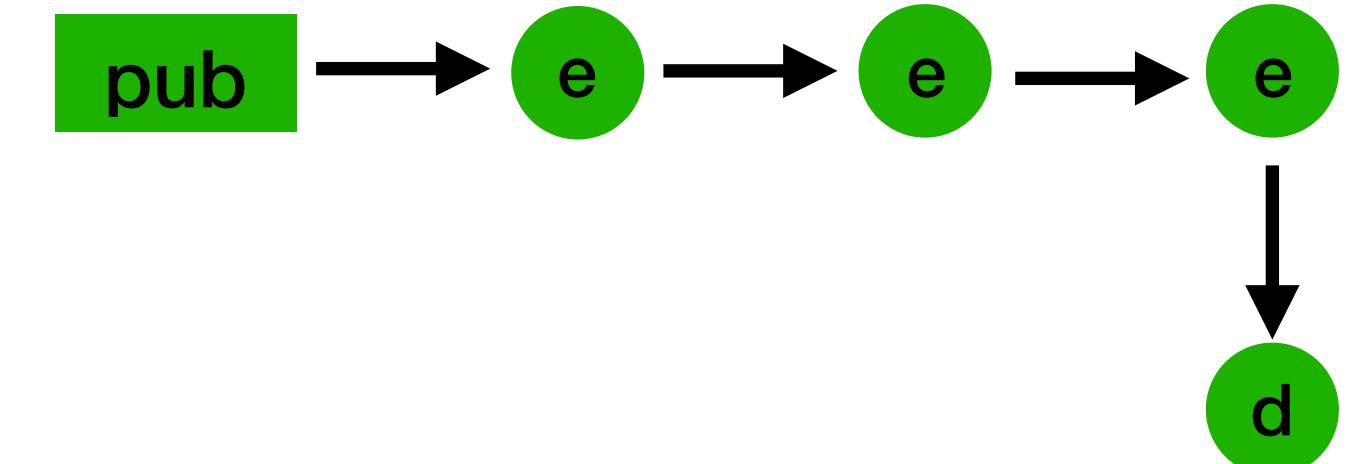
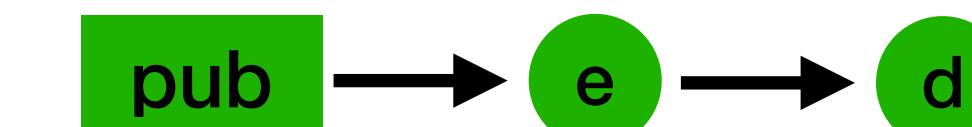
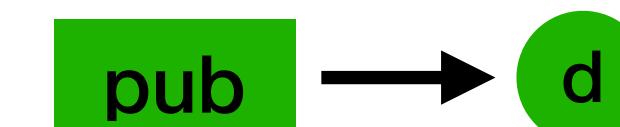
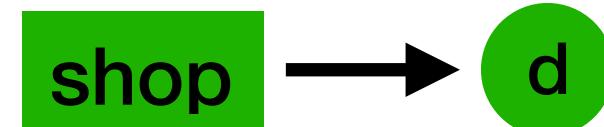
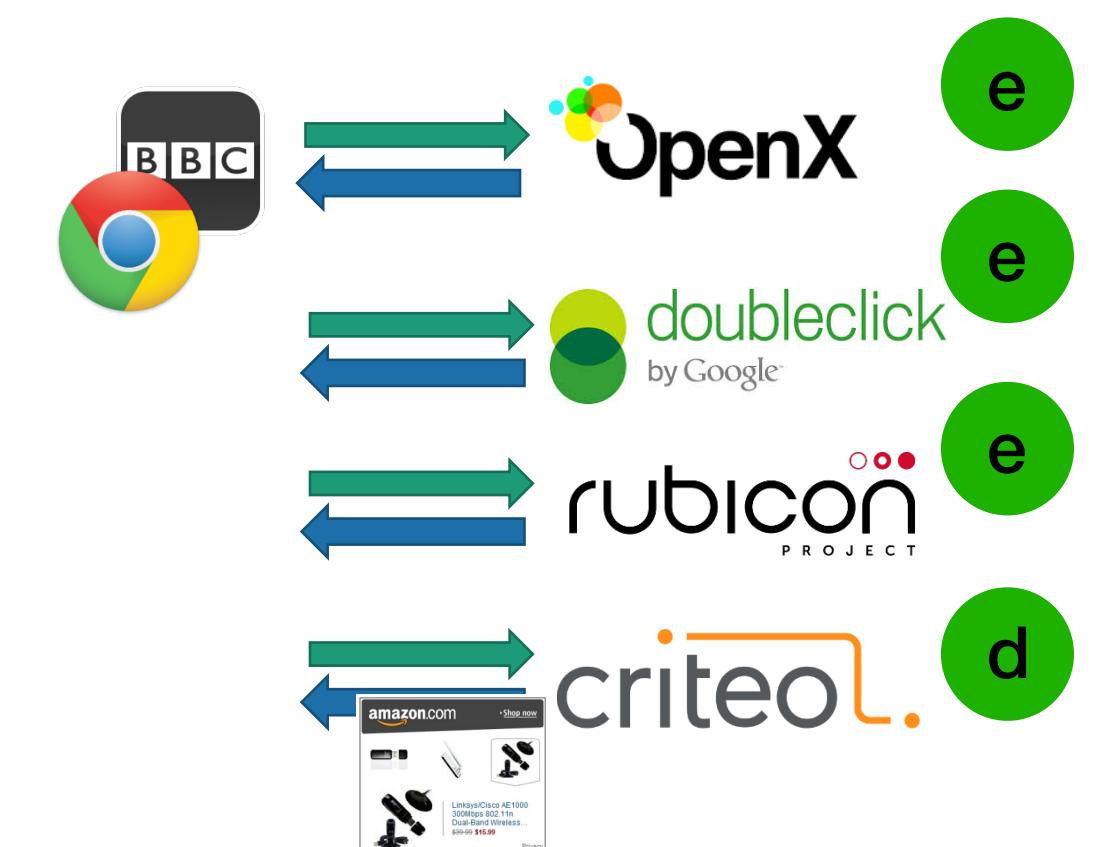
Shopper-side chain



Publisher-side chain



- Demand Side Platform (DSP)
- Exchange (Running Auction)



A Pattern for Chain Classification

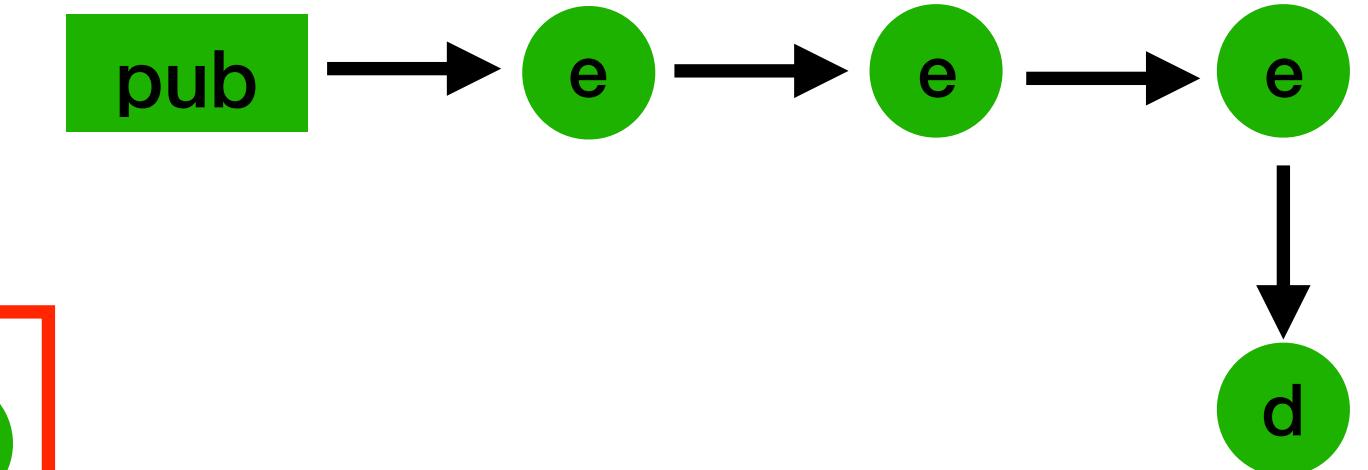
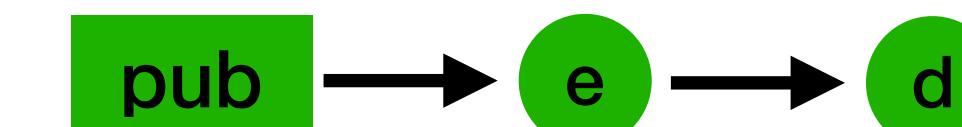
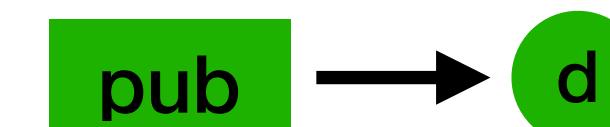
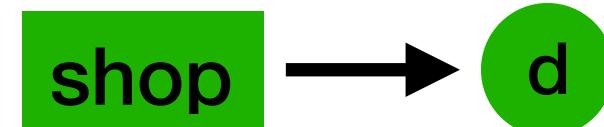
Shopper-side chain



Publisher-side chain



- Demand Side Platform (DSP)
- Exchange (Running Auction)

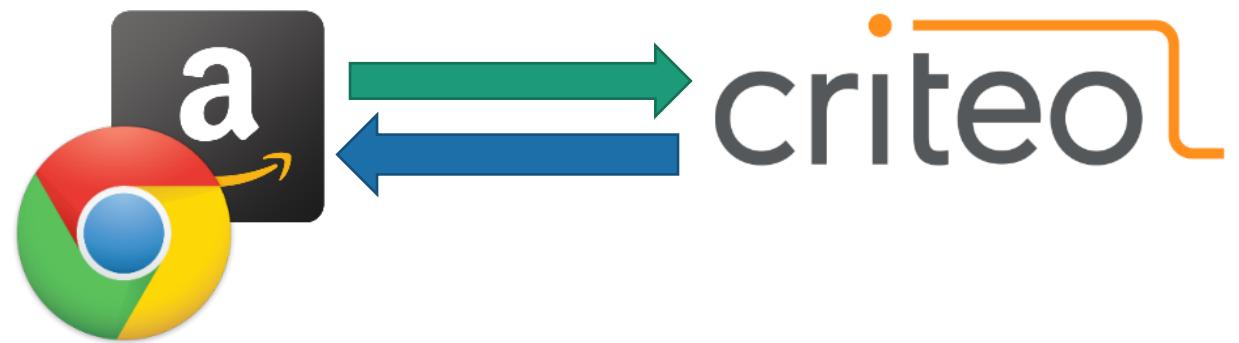


Regular Expression Like Rules

Direct (Trivial) Matching

Rule

Example



Shopper-side chain

Publisher-side chain

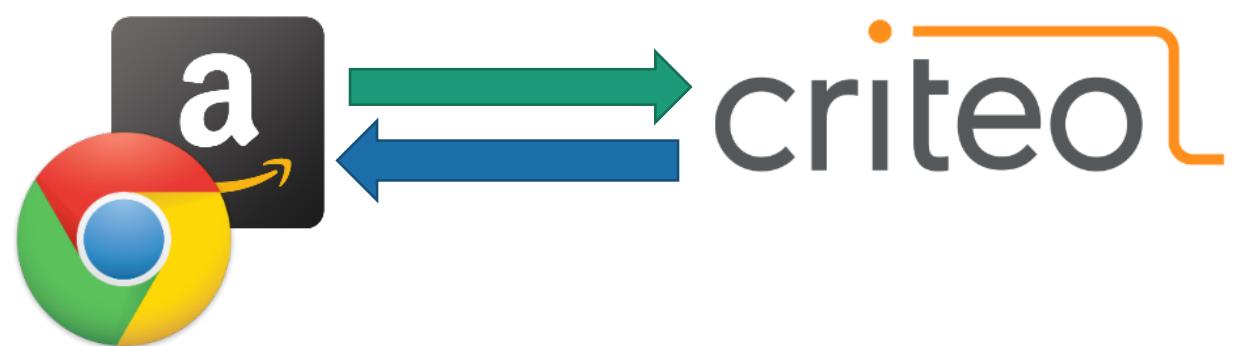


- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Direct (Trivial) Matching

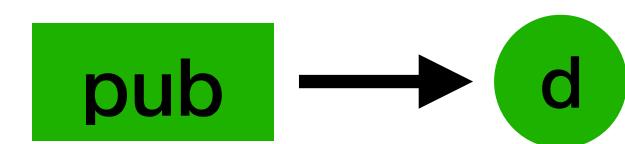
Rule

Example



Shopper-side chain

Publisher-side chain



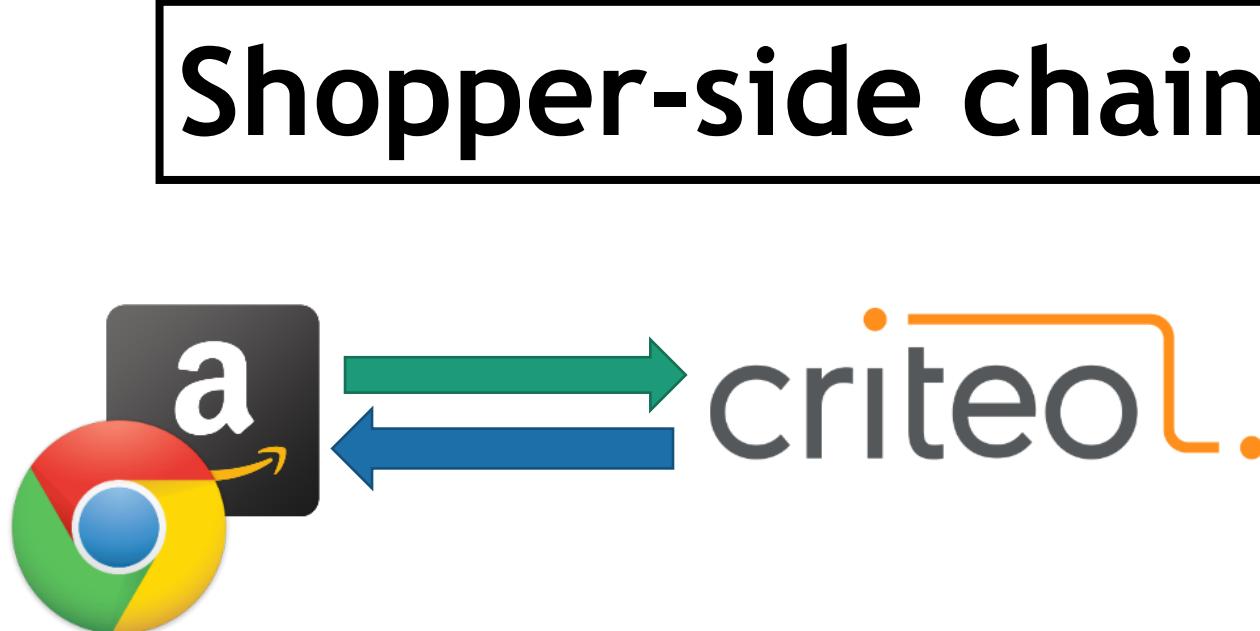
d is the DSP
that serves the
retarget

- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

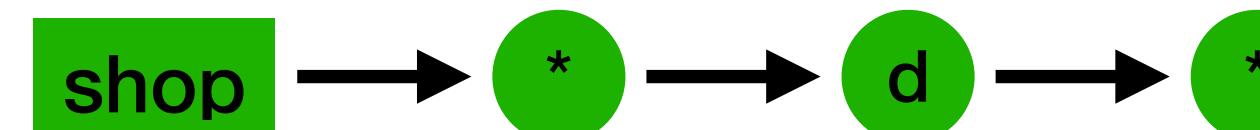
Direct (Trivial) Matching



Example

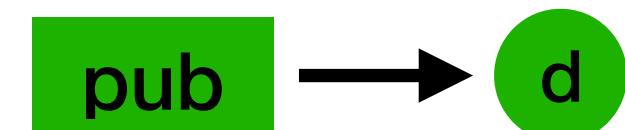


Rule



DSP d must observe us on the shopper side

... but other trackers may also appear

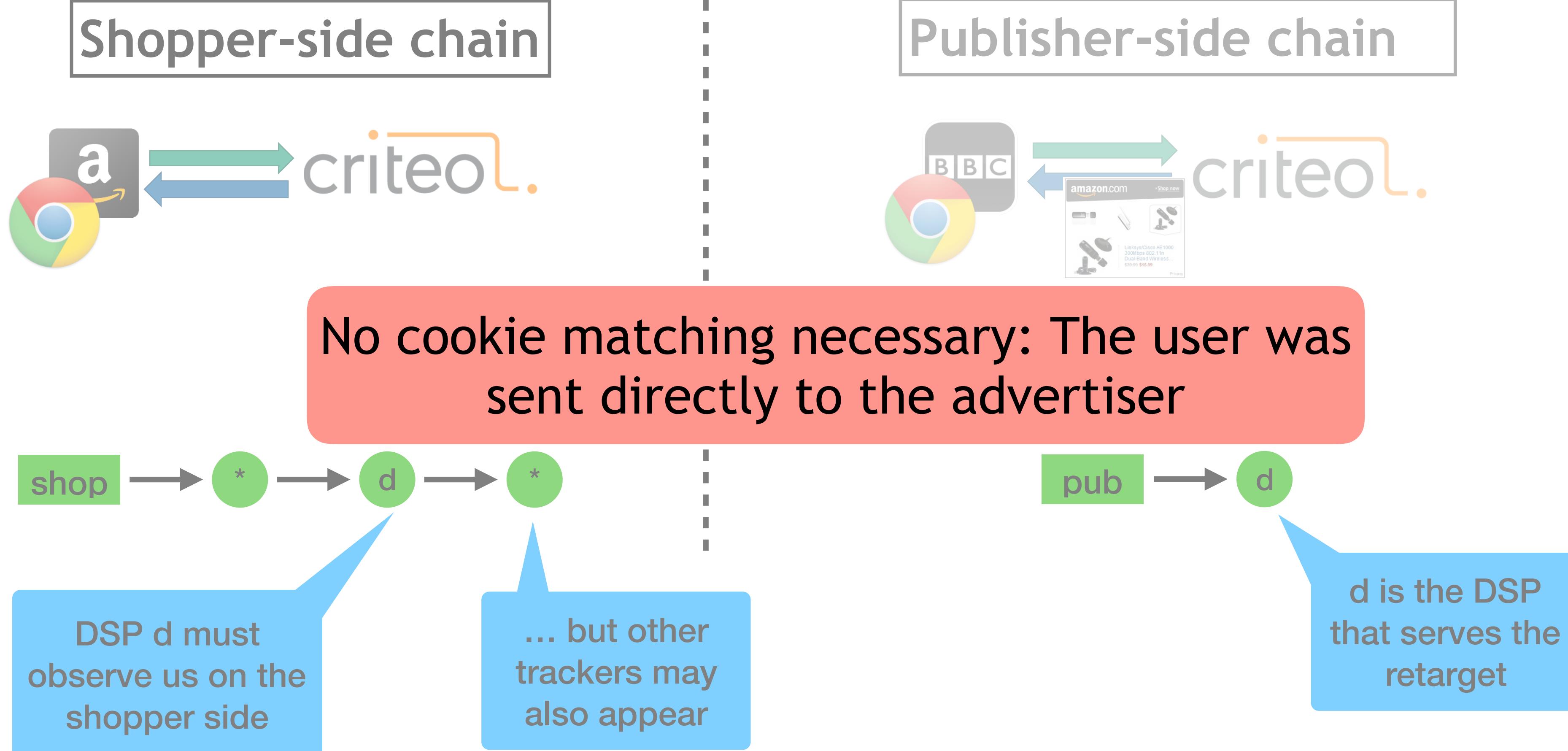


d is the DSP that serves the retarget

Direct (Trivial) Matching

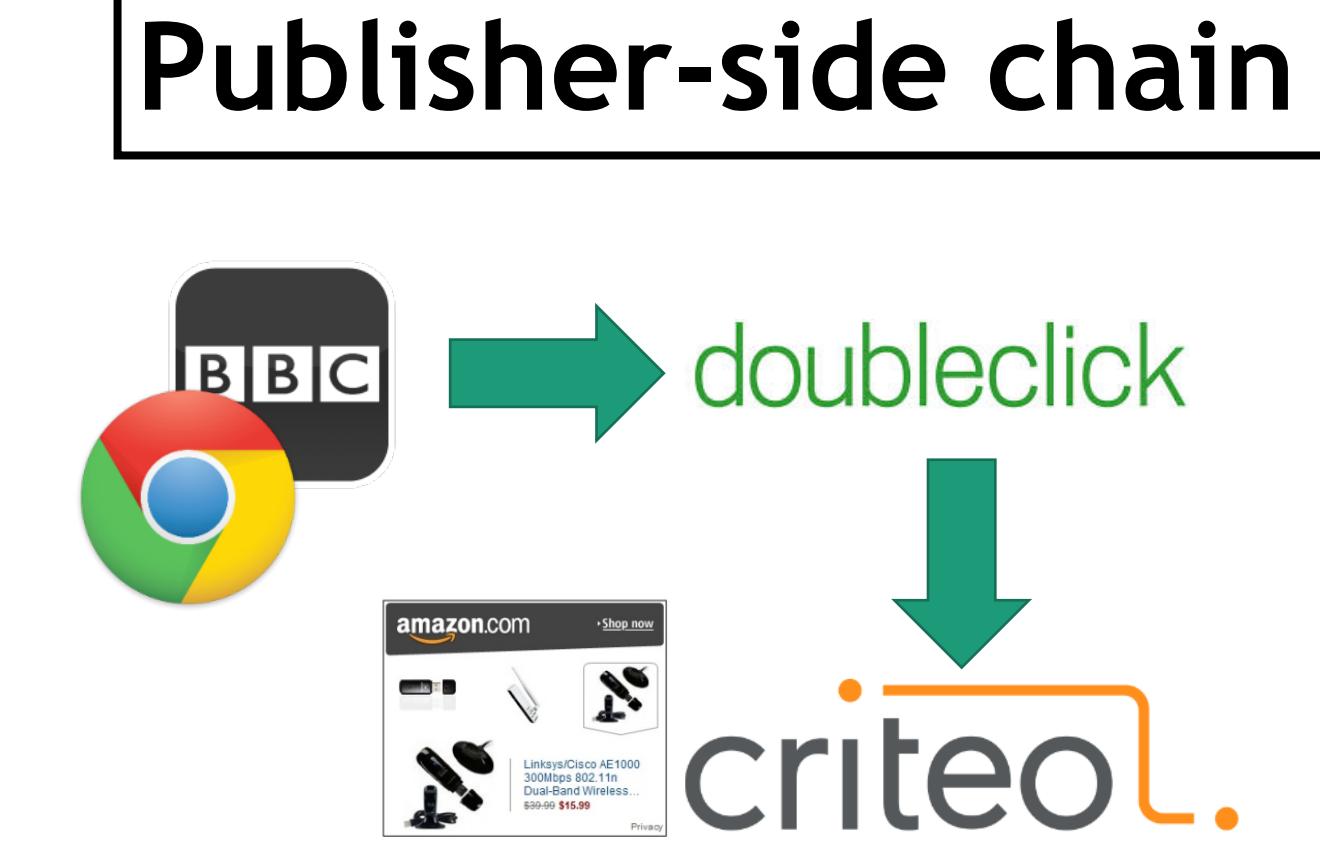
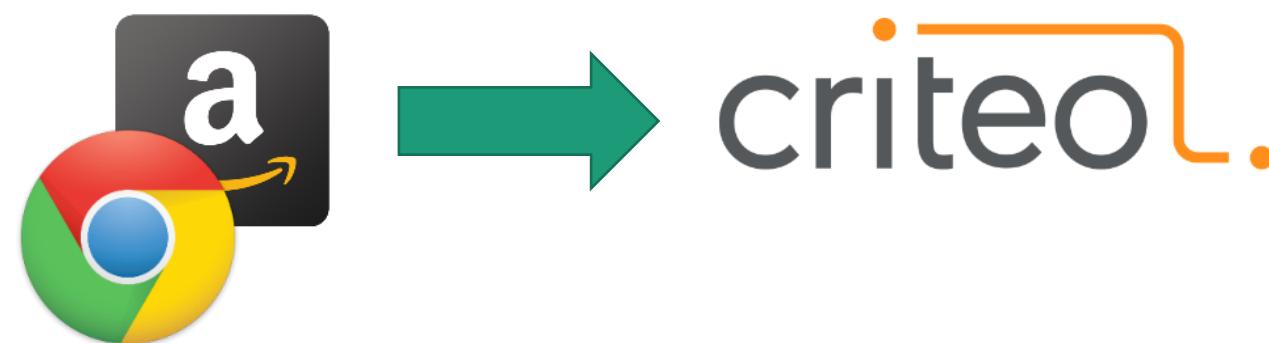
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Example



Cookie Matching

Example
Rule



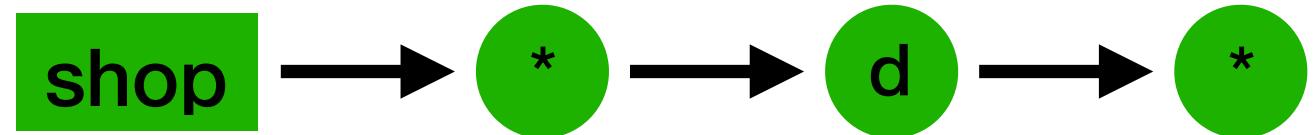
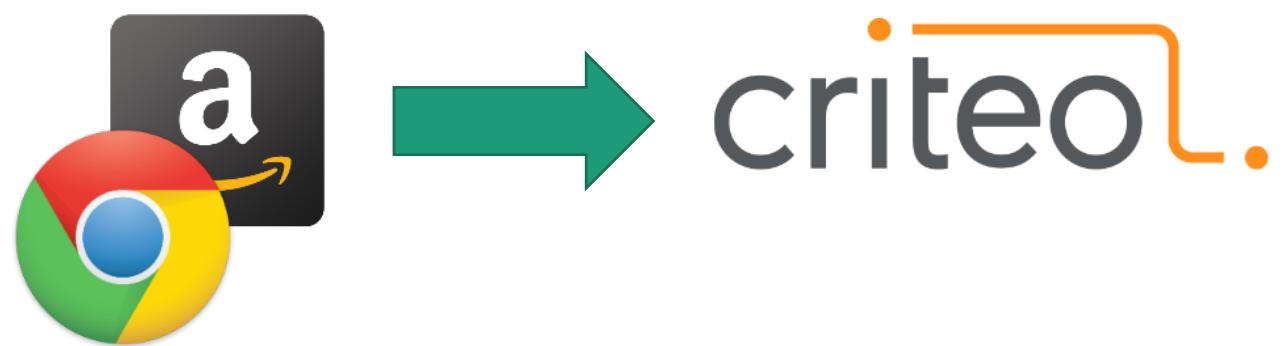
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Cookie Matching

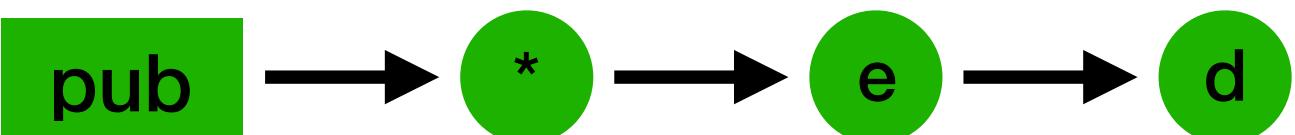
Example

Rule

Shopper-side chain



Publisher-side chain



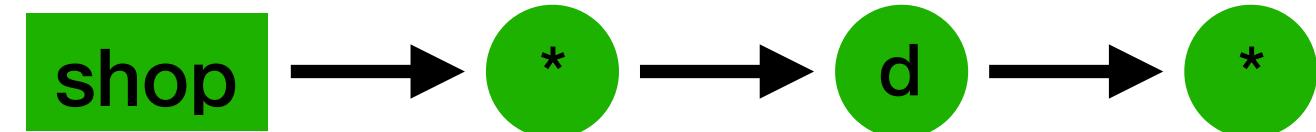
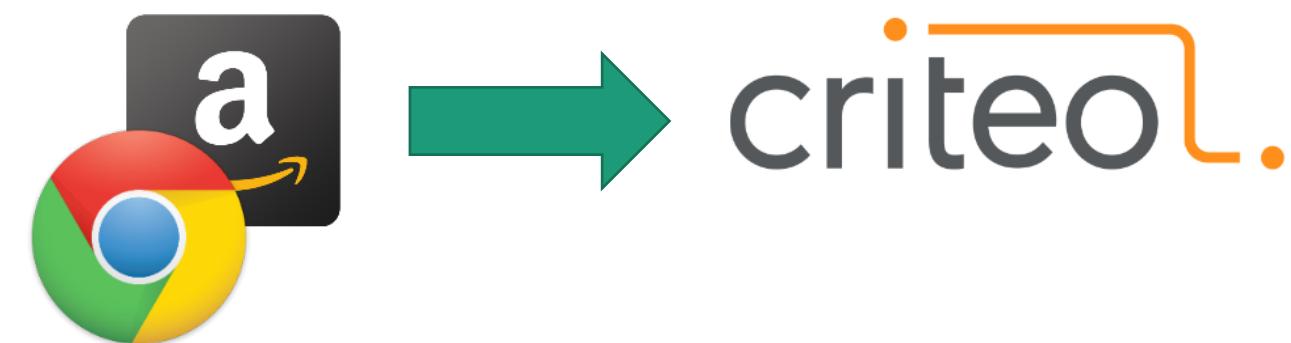
- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Cookie Matching

Example

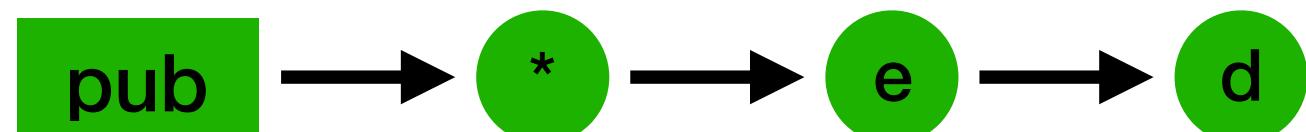
Rule

Shopper-side chain

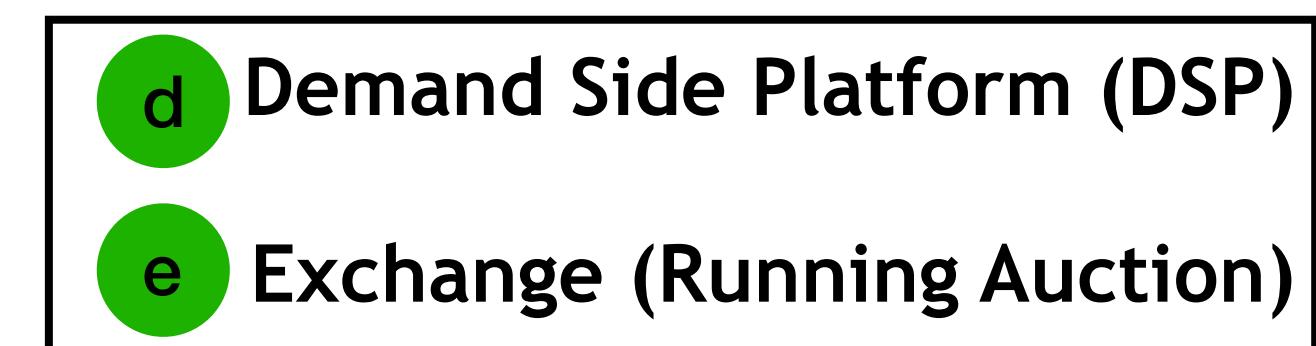


DSP d must observe us on the shopper side

Publisher-side chain



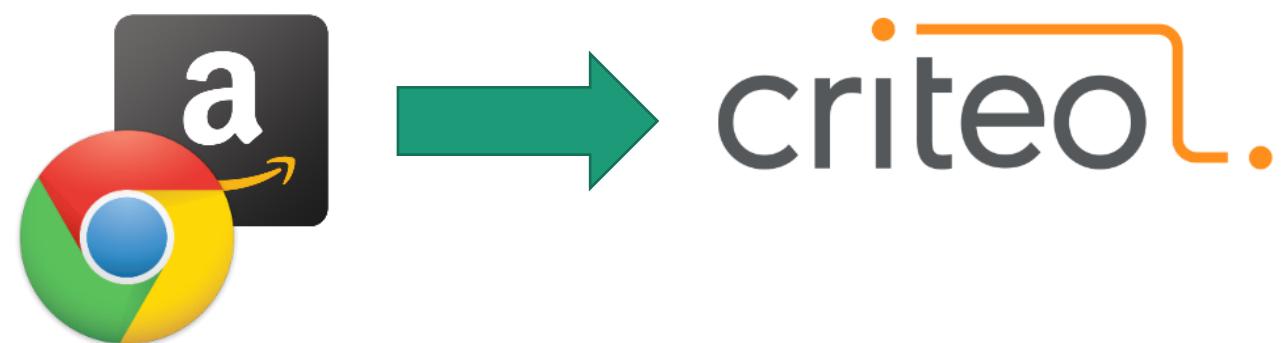
e precedes d, which implies an RTB auction



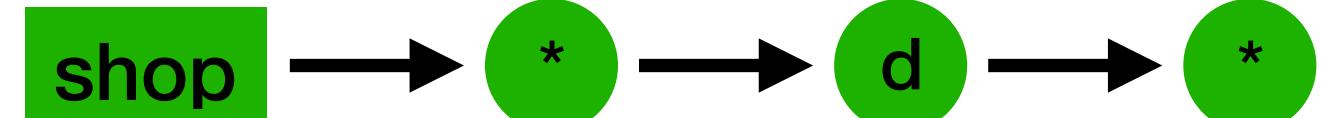
Cookie Matching

Example

Shopper-side chain

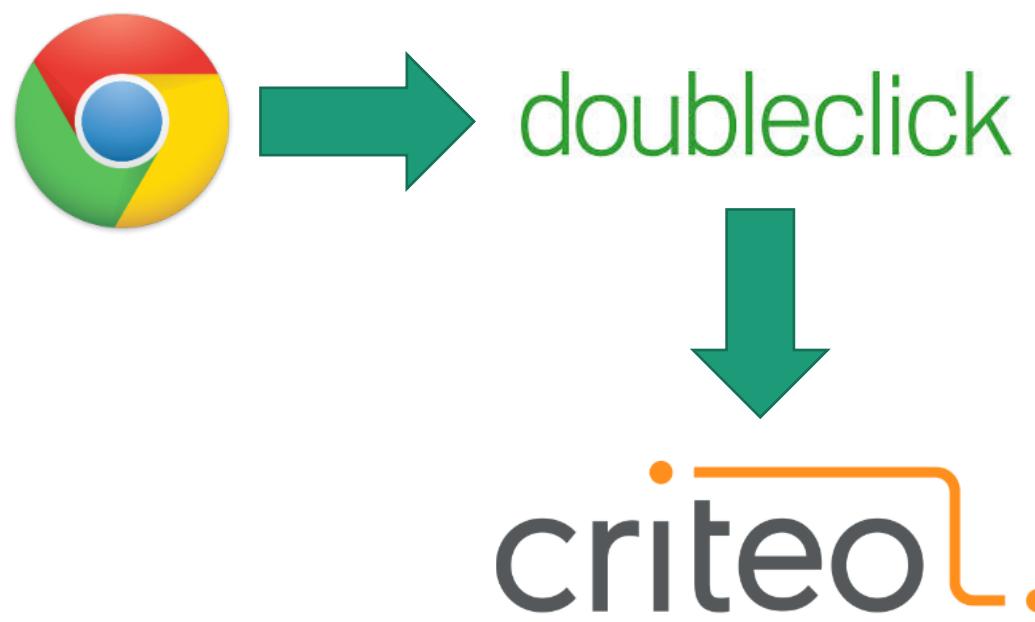


Rule



DSP d must observe us on the shopper side

Anywhere



Publisher-side chain



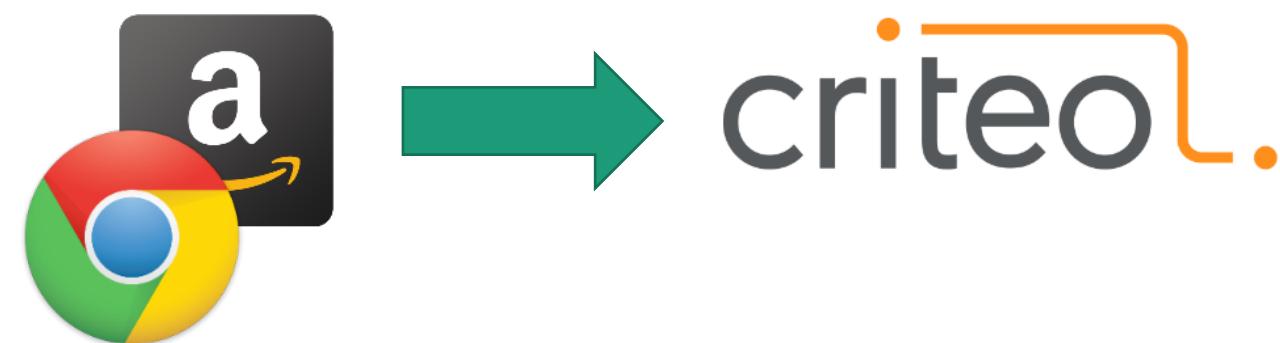
e precedes d, which implies an RTB auction

- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

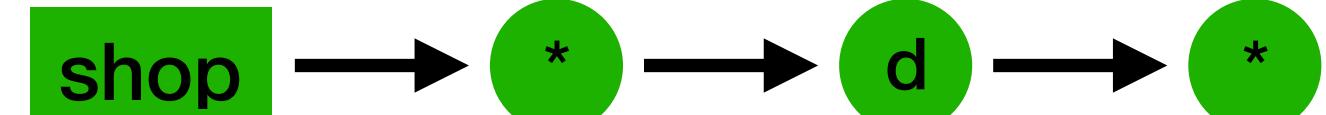
Cookie Matching

Example

Shopper-side chain

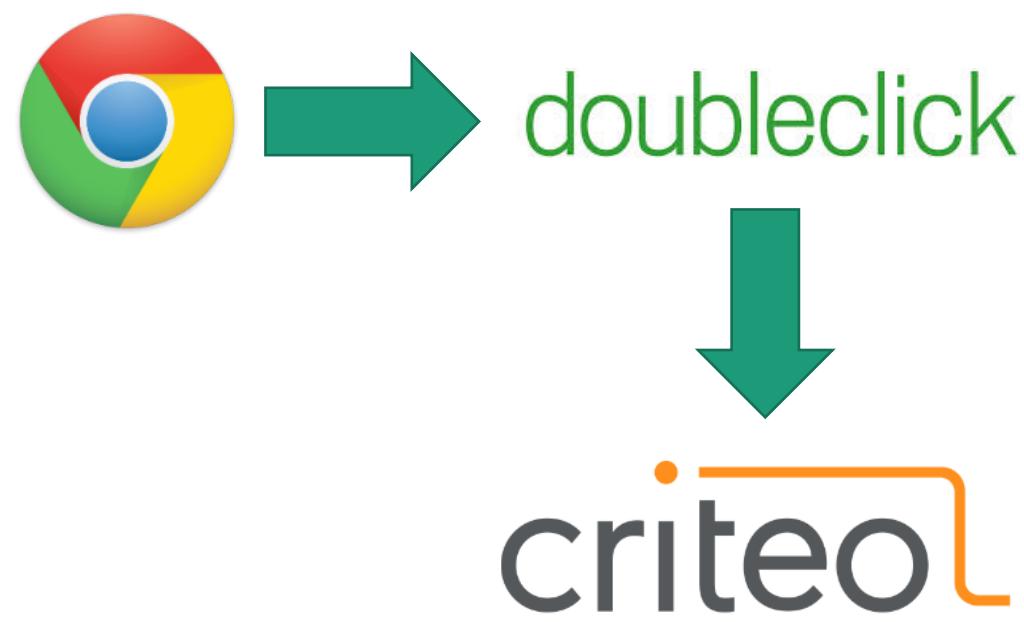


Rule

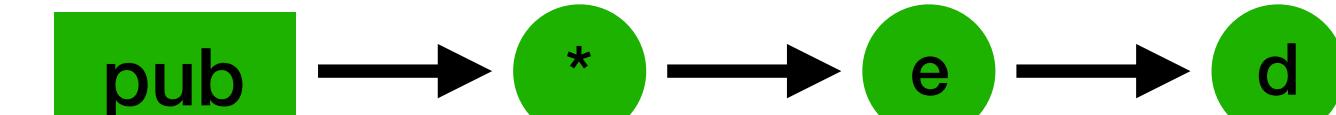


DSP d must observe us on the shopper side

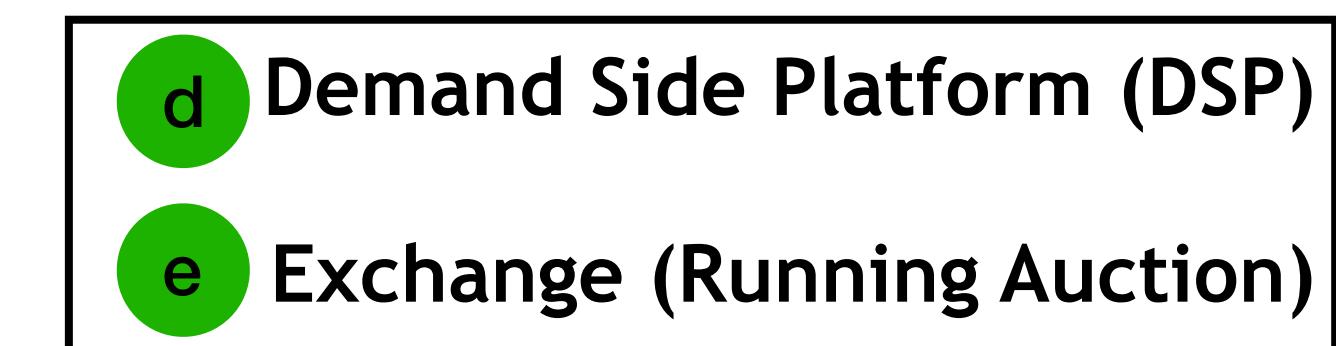
Anywhere



Publisher-side chain



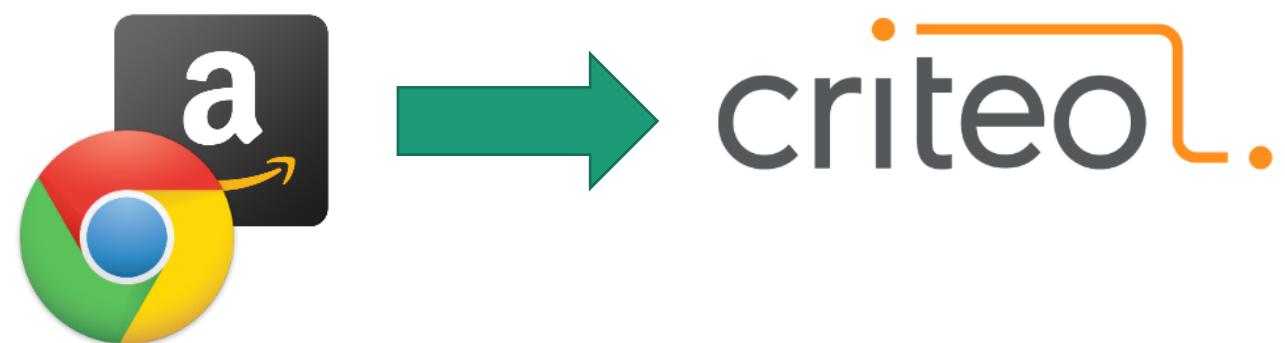
e precedes d, which implies an RTB auction



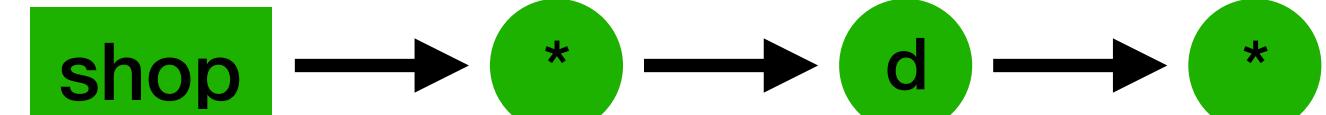
Cookie Matching

Example

Shopper-side chain

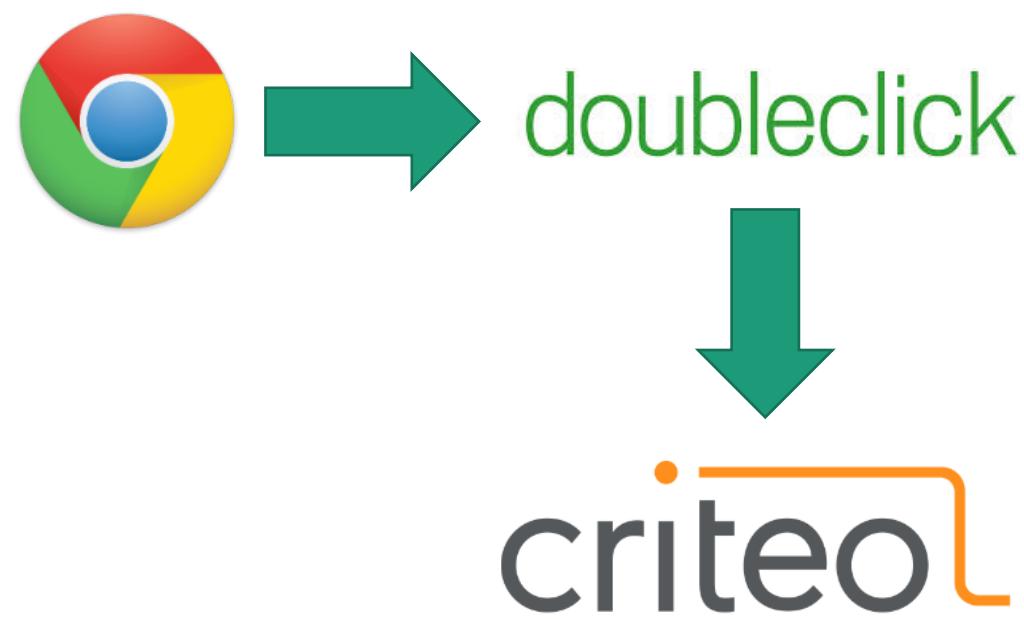


Rule



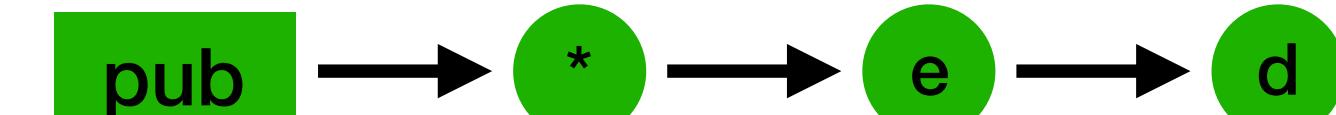
DSP d must observe us on the shopper side

Anywhere



Transition e—>d is where cookie match happens

Publisher-side chain



e precedes d, which implies an RTB auction

-  Demand Side Platform (DSP)
-  Exchange (Running Auction)

Cookie Matching

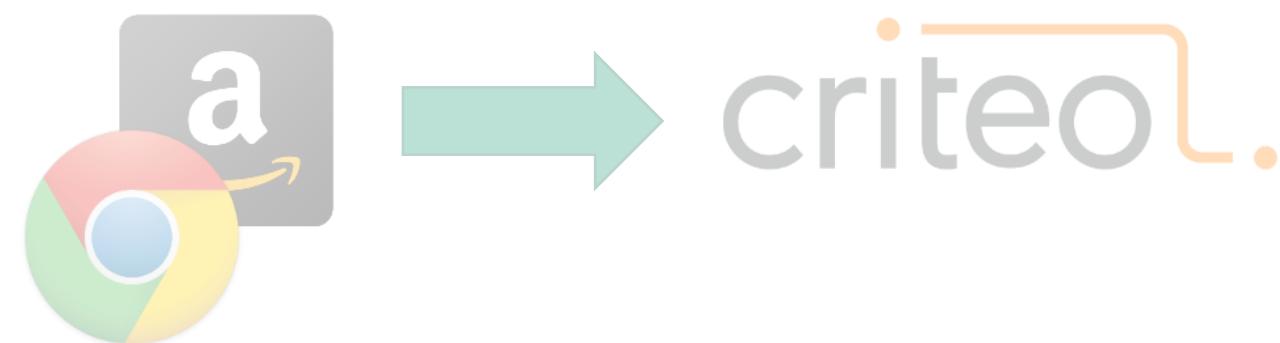
Example

Rule

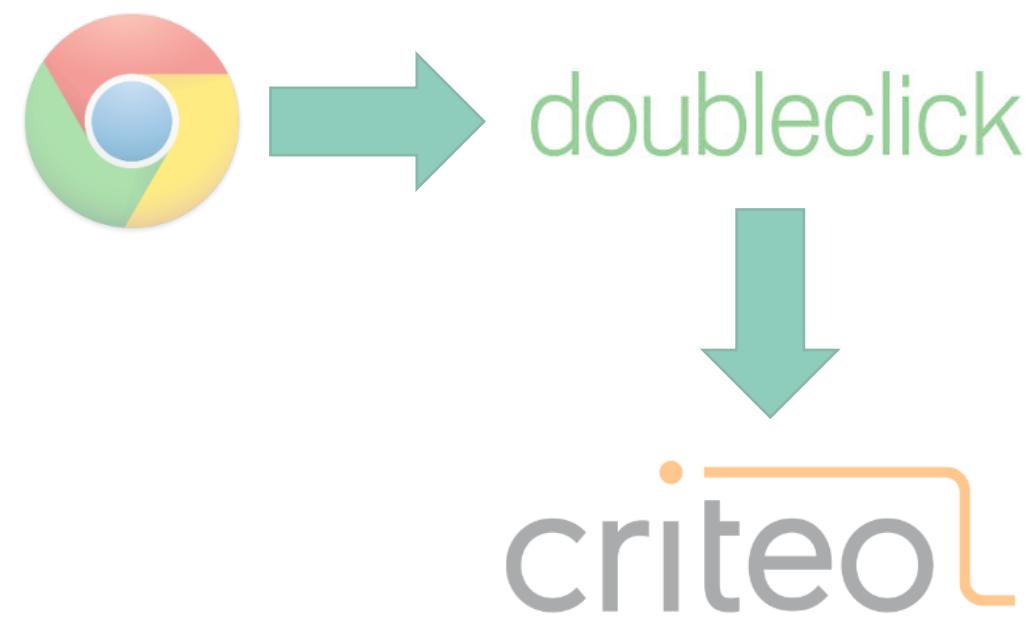
e → d shows that e synched identifiers with d before RTB auction

- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Shopper-side chain



Anywhere



Publisher-side chain



DSP d must observe us on the shopper side



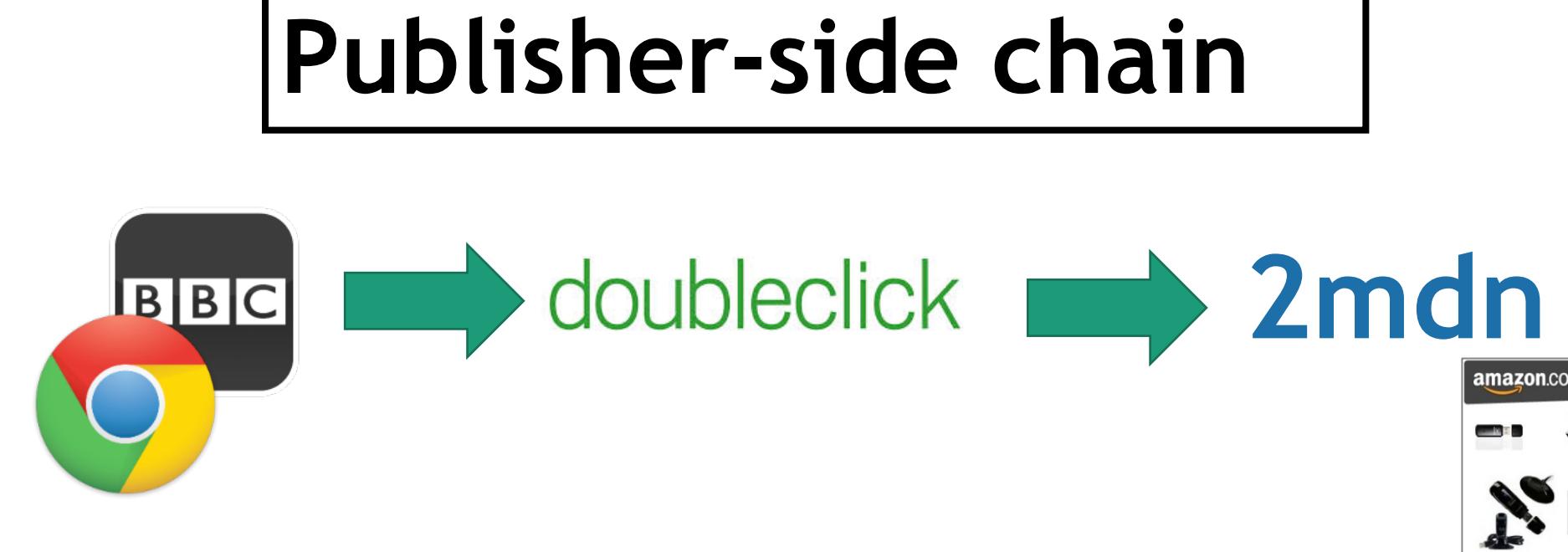
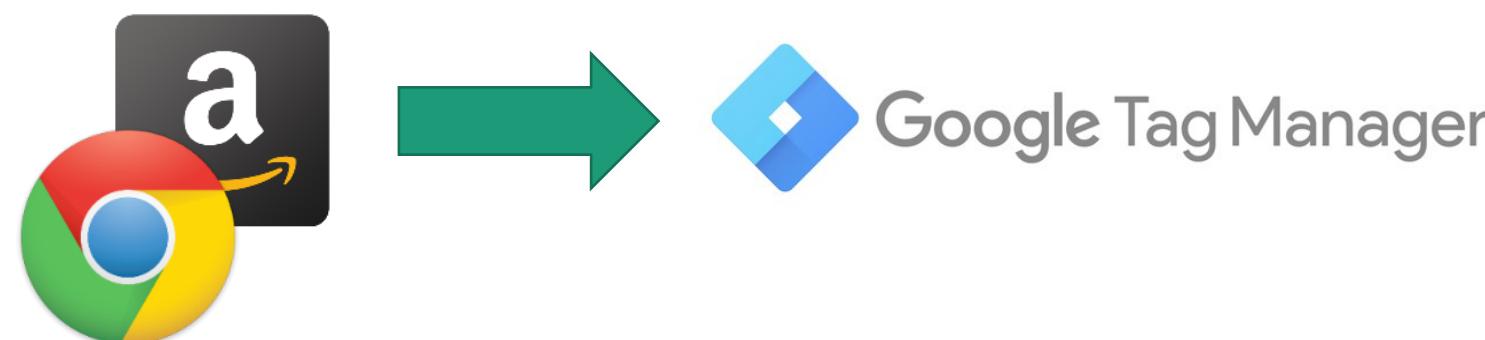
Transition e → d is where cookie match happens



e precedes d, which implies an RTB auction

Latent (Server-Side) Matching

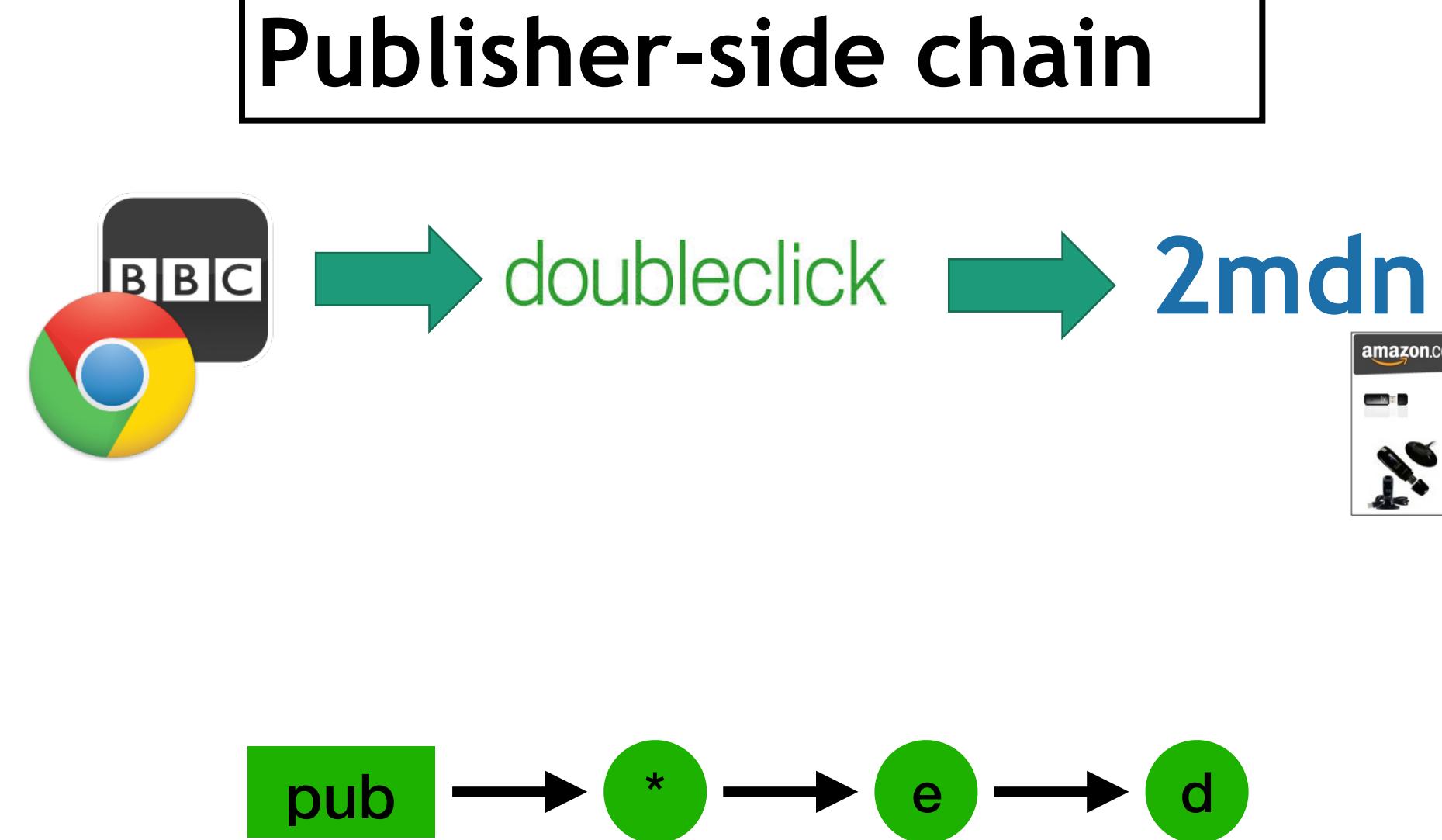
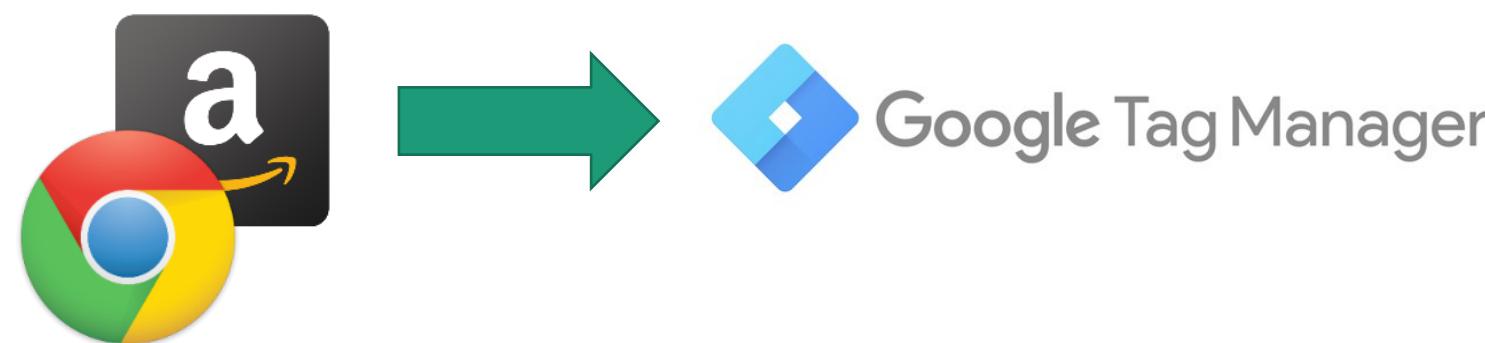
Rule Example



- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Latent (Server-Side) Matching

Rule Example

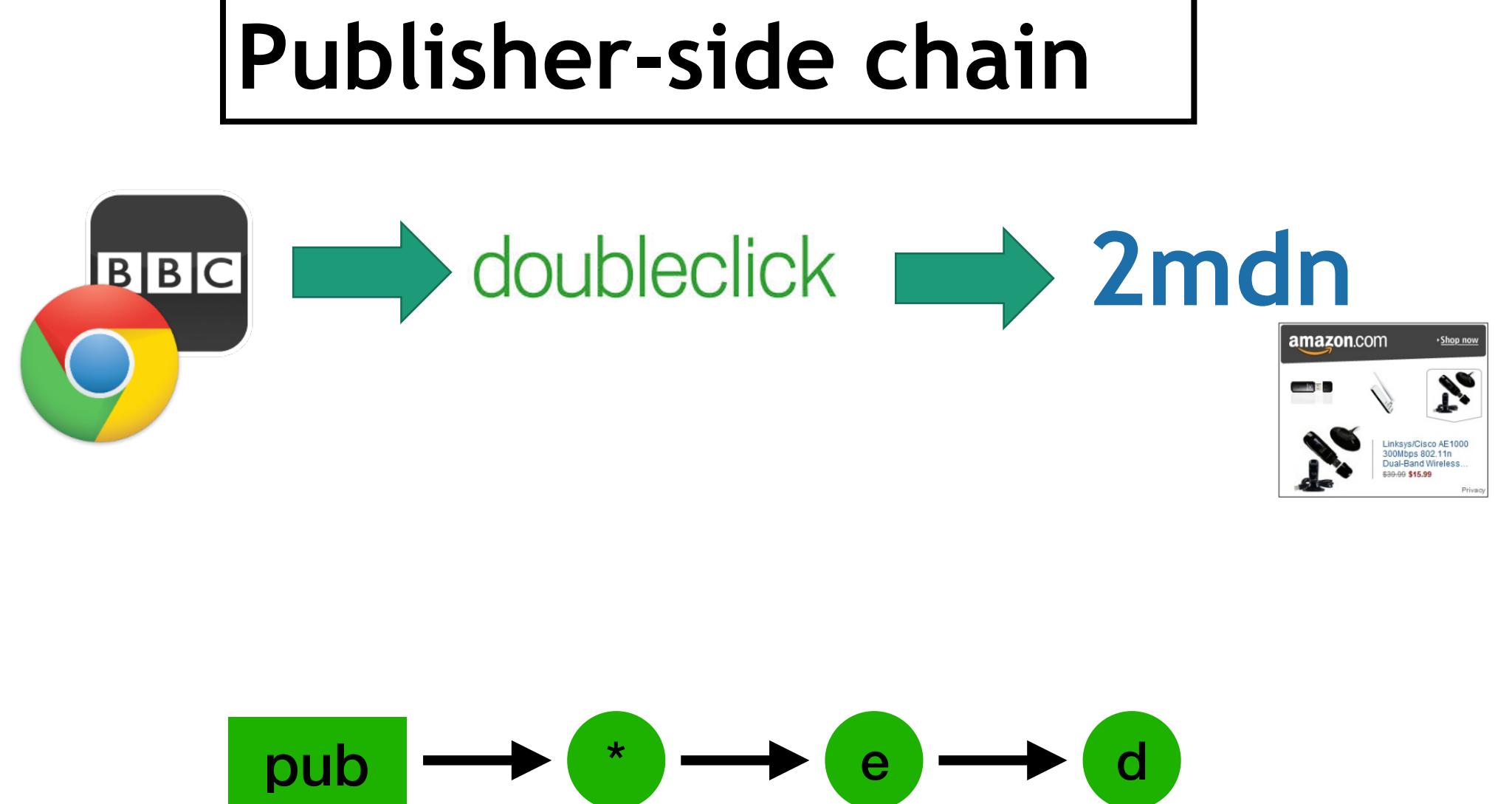
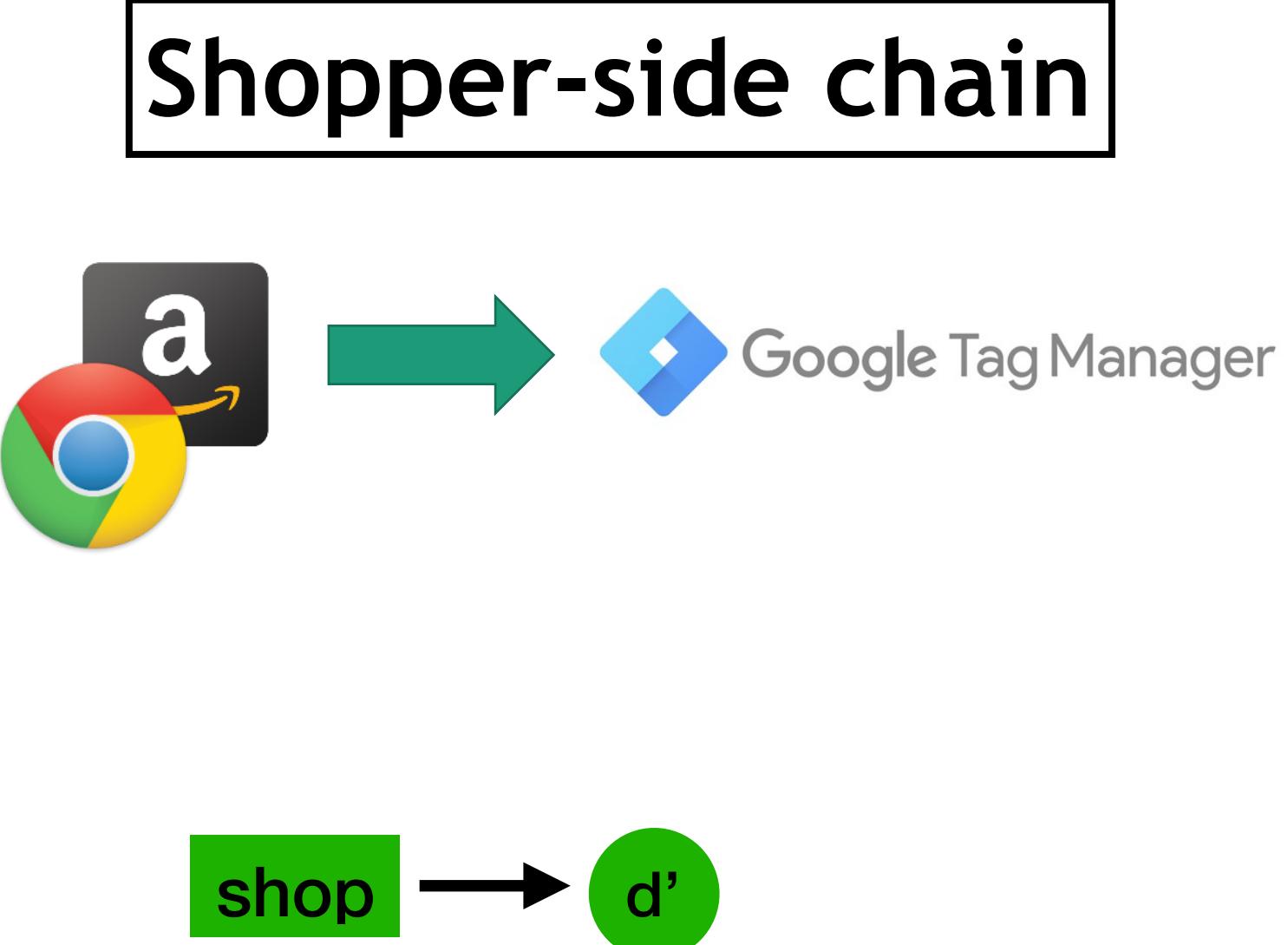


- d Demand Side Platform (DSP)
- e Exchange (Running Auction)

Latent (Server-Side) Matching



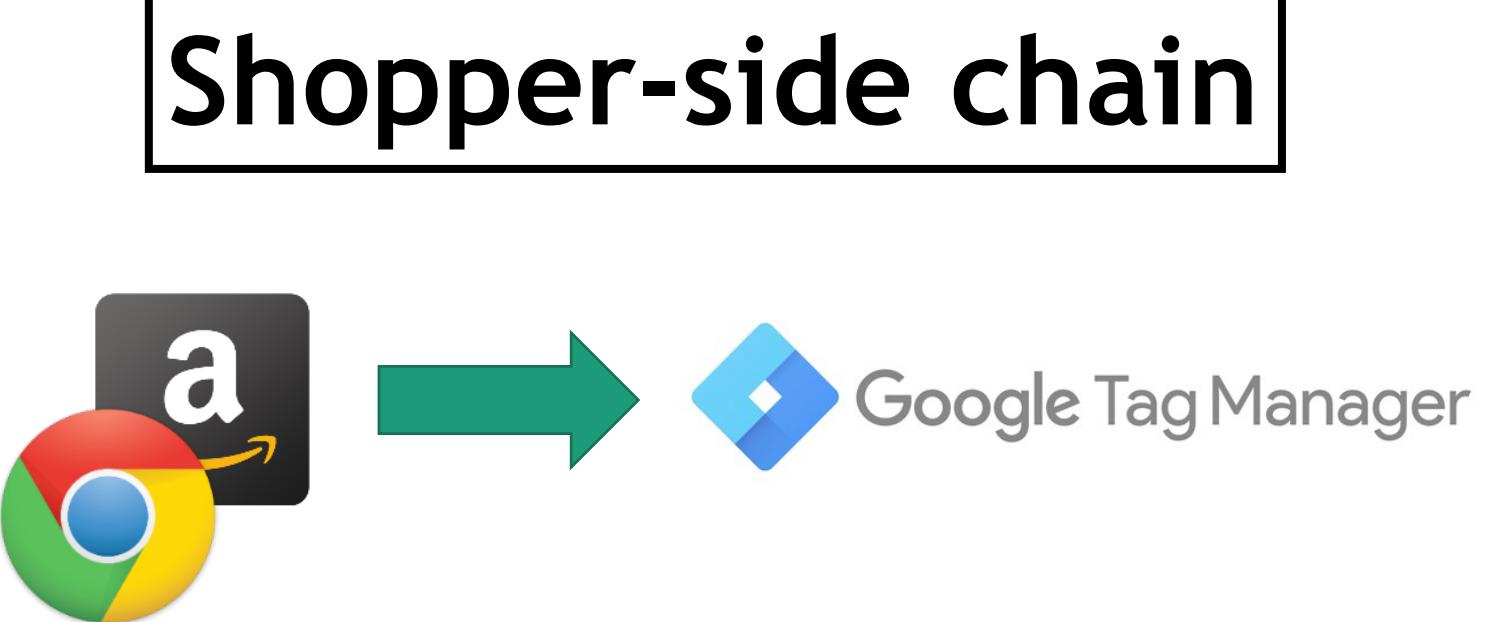
Example
Rule



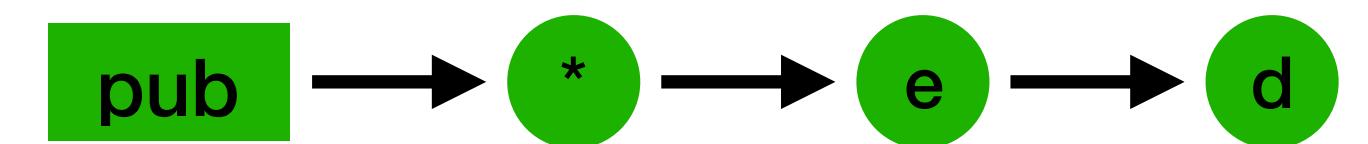
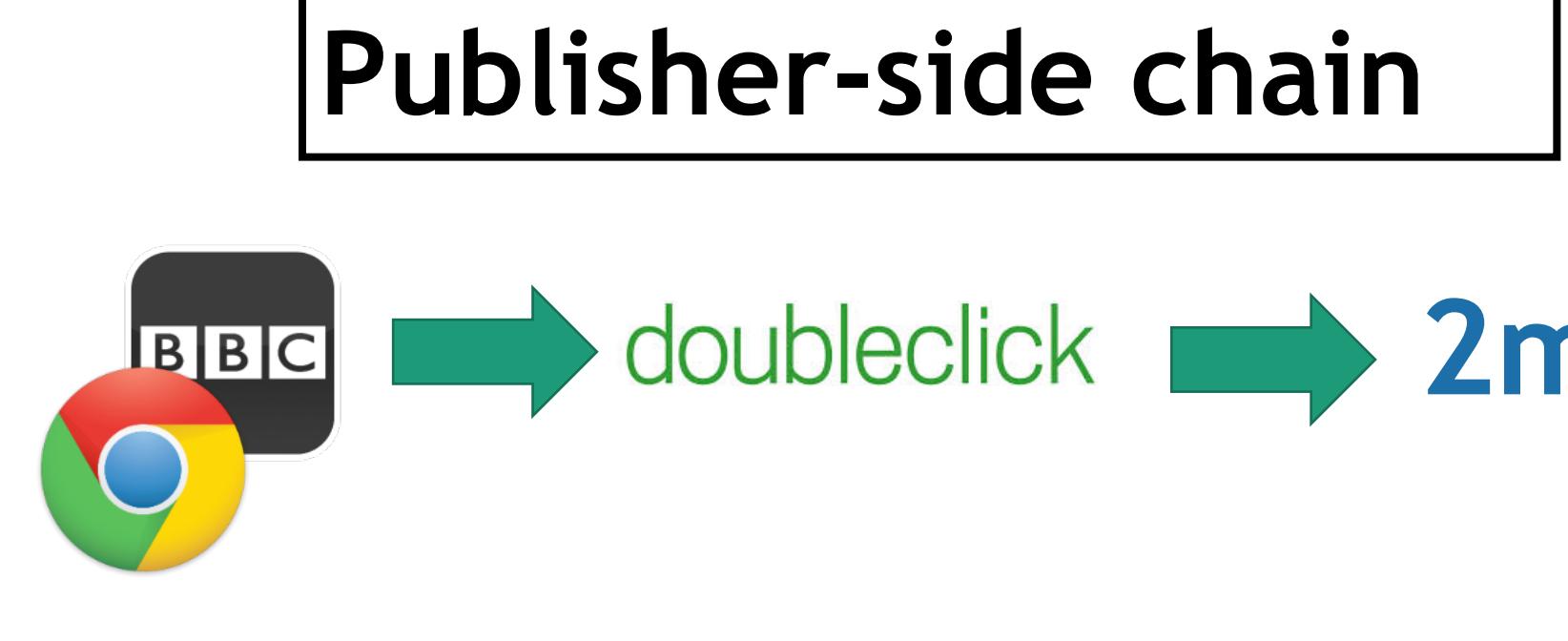
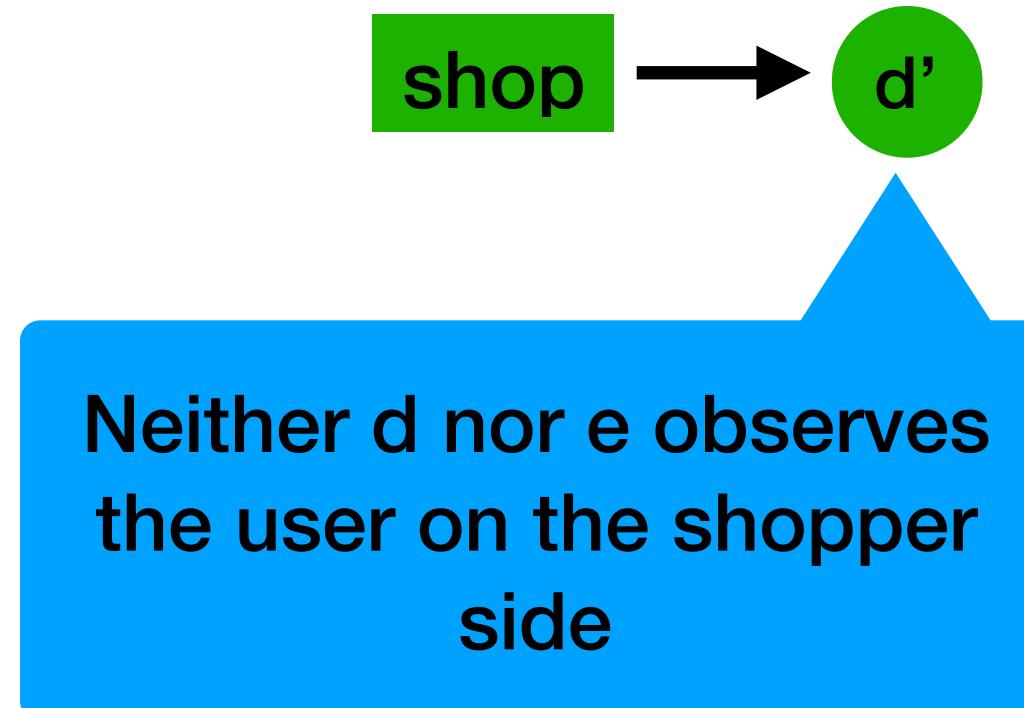
Latent (Server-Side) Matching



Example



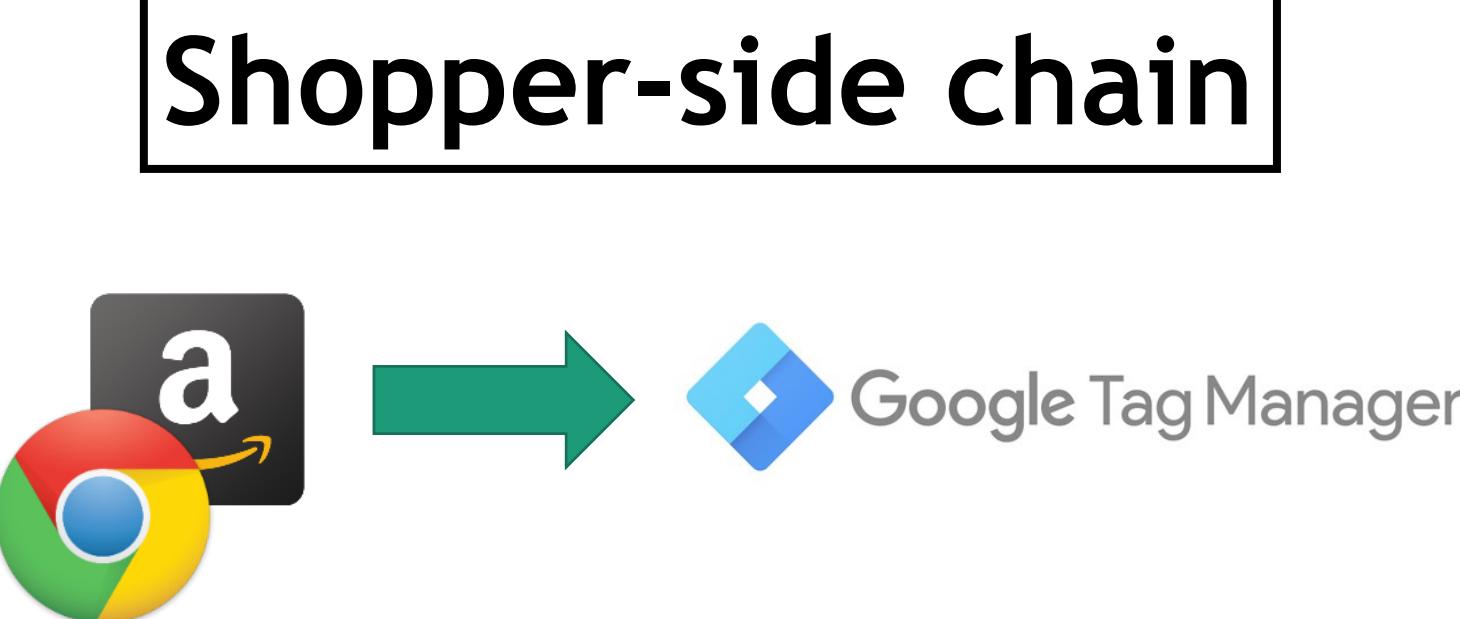
Rule



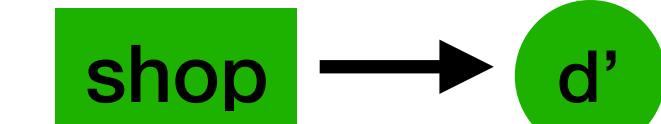
Latent (Server-Side) Matching



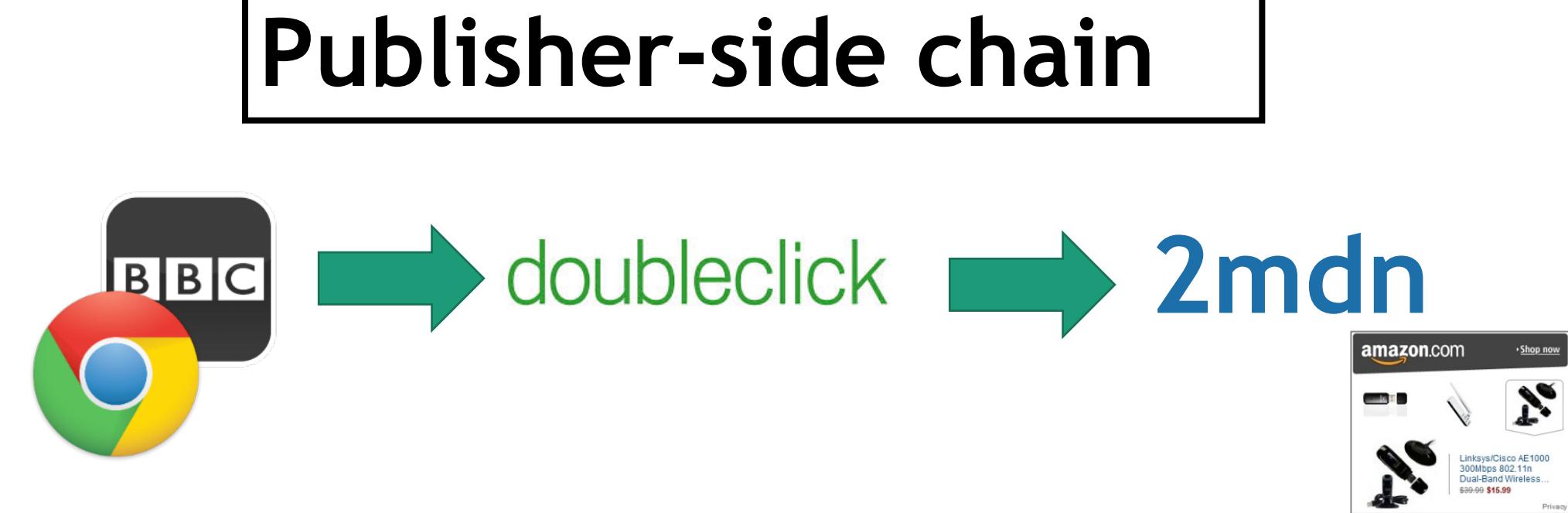
Example



Rule



Neither d nor e observes the user on the shopper side



d must receive information from some shopper-side tracker

d and d' must have shared user identifiers on the server-side

Overall Categorization of Chains

Type	Raw Chains		Clustered Chains	
	Chains	%	Chains	%
Direct (Trivial) Match	1770	5	8449	24
Cookie Match	25049	71	25873	73
Latent Match	5362	15	343	1
No Match	775	2	183	1

Overall Categorization of Chains

Type	Raw Chains		Clustered Chains	
	Chains	%	Chains	%
Direct (Trivial) Match	1770	5	8449	24
Cookie Match	25049	71	25873	73
Latent Match	5362	15	343	1
No Match	775	2	183	1

Latent matches essentially disappear

- The vast majority of these chains involved Google
- Suggests that Google shares tracking data across their services

Who are the Cookie Matching Partners?

Participant 1		Participant 2	#Chains	#Ads
criteo	↔	googlesyndication	9090	1887
criteo	↔	doubleclick	3610	1144
criteo	↔	adnxss	3263	1066
criteo	↔	rubiconproject	1586	749
criteo	↔	servedbyopenx	707	460
doubleclick	↔	steelhousemedia	362	27
mathtag	↔	mediaforge	360	124
netmng	↔	scene7	267	119
googlesyndication	↔	adsrvr	107	29
rubiconproject	↔	steelhousemedia	86	30
googlesyndication	↔	steelhousemedia	47	22
adtechus	→	adacado	36	18
atwola	→	adacado	32	6
adroll	↔	adnxss	31	8

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads
criteo	← → googlesyndication	9090	1887
criteo	← → doubleclick	3610	1144
criteo	← → adnxss	3263	1066
criteo	← → rubiconproject	1586	749
criteo	← → servedbyopenx	707	460
doubleclick	← → steelhousemedia	362	27
mathtag	← → mediaforge	360	124
netmng	← → scene7	267	119
googlesyndication	← → adsrvr	107	29
rubiconproject	← → steelhousemedia	86	30
googlesyndication	← → steelhousemedia	47	22
adtechus	→ adacado	36	18
atwola	→ adacado	32	6
adroll	← → adnxss	31	8

Direction in which
matching happens

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads
criteo	← →	googlesyndication	9090 1887
criteo	← →	doubleclick	3610 1144
criteo	← →	adnxss	3263 1066
criteo	← →	rubiconproject	1586 749
criteo	← →	servedbyopenx	707 460
doubleclick	← →	steelhousemedia	362 27
mathtag	← →	mediaforge	360 124
netmng	← →	scene7	267 119
googlesyndication	← →	adsrvr	107 29
rubiconproject	← →	steelhousemedia	86 30
googlesyndication	← →	steelhousemedia	47 22
adtechus	→	adacado	36 18
atwola	→	adacado	32 6
adroll	← →	adnxss	31 8

Direction in which
matching happens

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads	Heuristics (Prior Work)
criteo	↔ ↔	googlesyndication	9090	1887
criteo	↔ ↔	doubleclick	3610	1144 → E, P ← DC, P
criteo	↔ ↔	adnxs	3263	1066 ← → E, P
criteo	↔ ↔	rubiconproject	1586	749 ← → E, P
criteo	↔ ↔	servedbyopenx	707	460 ← → P
doubleclick	↔ ↔	steelhousemedia	362	27 → P ← E, P
mathtag	↔ ↔	mediaforge	360	124 ← → E, P
netmng	↔ ↔	scene7	267	119 → E ← ?
googlesyndication	↔ ↔	adsrvr	107	29 ← → P
rubiconproject	↔ ↔	steelhousemedia	86	30 ← → E
googlesyndication	↔ ↔	steelhousemedia	47	22 ?
adtechus	→	adacado	36	18 ?
atwola	→	adacado	32	6 ?
adroll	↔ ↔	adnxs	31	8 ?

Heuristics Key (used by prior work)

E - exact identifier value shared

P - special URL parameters
e.g. cookie_sync, user_sync

DC - DoubleClick URL parameters
e.g. google_gid, google_nid

? - Match missed by heuristics

Direction in which matching happens

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads	Heuristics (Prior Work)
criteo	↔ ↔	googlesyndication	9090	1887 ↔ → P
criteo	↔ ↔	doubleclick	3610	1144 → E, P ← DC, P
criteo	↔ ↔	adnxns	3263	1066 ↔ → E, P
criteo	↔ ↔	rubiconproject	1586	749 ↔ → E, P
criteo	↔ ↔	servedbyopenx	707	460 ↔ → P
doubleclick	↔ ↔	steelhousemedia	362	27 → P ← E, P
mathtag	↔ ↔	mediaforge	360	124 ↔ → E, P
netmng	↔ ↔	scene7	267	119 → E ← ?
googlesyndication	↔ ↔	adsrvr	107	29 ↔ → P
rubiconproject	↔ ↔	steelhousemedia	86	30 ↔ → E
googlesyndication	↔ ↔	steelhousemedia	47	22 ?
adtechus	→	adacado	36	18 ?
atwola	→	adacado	32	6 ?
adroll	↔ ↔	adnxns	31	8 ?

Heuristics Key (used by prior work)

E - exact identifier value shared

P - special URL parameters
e.g. cookie_sync, user_sync

DC - DoubleClick URL parameters
e.g. google_gid, google_nid

? - Match missed by heuristics

Direction in which
matching happens

Who are the Cookie Matching Partners?

Participant 1	Participant 2	#Chains	#Ads	Heuristics (Prior Work)
criteo	↔ ↔	googlesyndication	9090	1887
criteo	↔ ↔	doubleclick	3610	1144 → E, P ← DC, P
criteo	↔ ↔	adnxns	3263	1066 ← → E, P
criteo	↔ ↔	rubiconproject	1586	749 ← → E, P
criteo	↔ ↔	servedbyopenx	707	460 ← → P
doubleclick	↔ ↔	steelhousemedia	362	27 → P ← E, P
mathtag	↔ ↔	mediaforge	360	124 ← → E, P
netmng	↔ ↔	scene7	267	119 → E ← ?
googlesyndication	↔ ↔	adsrvr	107	29 ← → P
rubiconproject	↔ ↔	steelhousemedia	86	30 ← → E
googlesyndication	↔ ↔	steelhousemedia	47	22 ?
adtechus	→	adacado	36	18 ?
atwola	→	adacado	32	6 ?
adroll	↔ ↔	adnxns	31	8 ?

Direction in which matching happens

Heuristics Key (used by prior work)

E - exact identifier value shared

P - special URL parameters
e.g. cookie_sync, user_sync

DC - DoubleClick URL parameters
e.g. google_gid, google_nid

? - Match missed by heuristics

31% of cookie matching partners were missed by prior heuristics

Some High-level Properties of the Graph

Some High-level Properties of the Graph

- The graph is very dense.

Some High-level Properties of the Graph

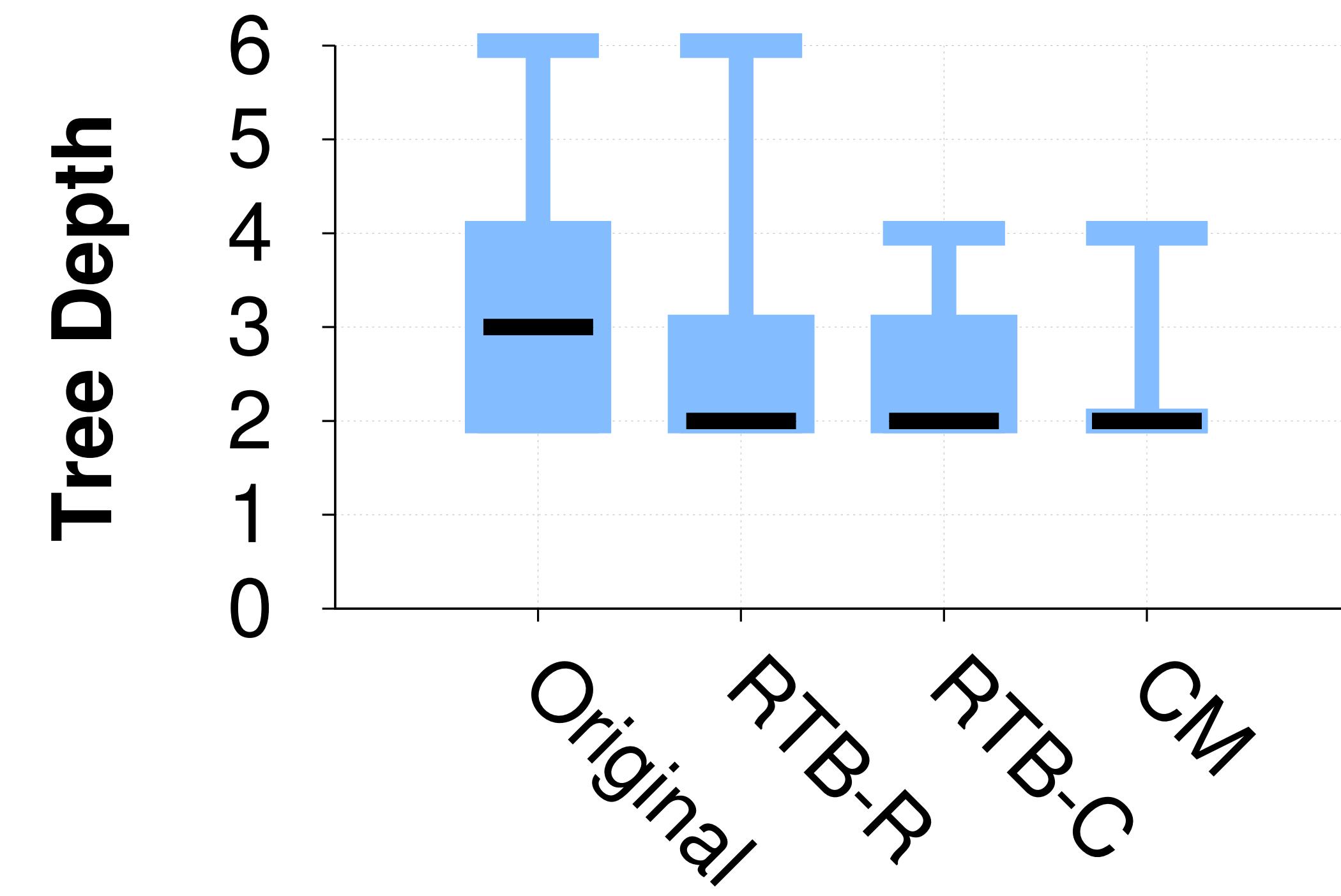
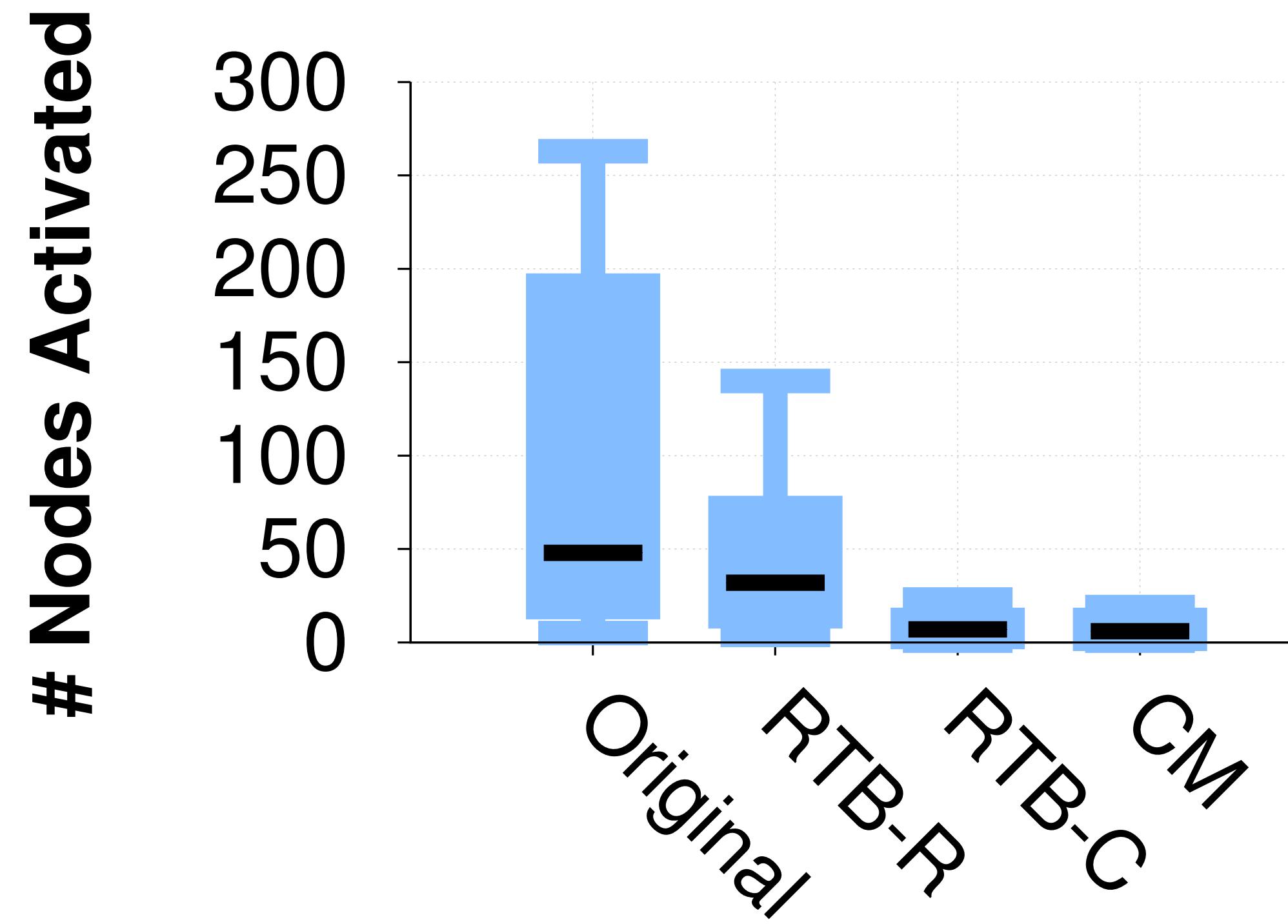
- The graph is very dense.
- No distinct communities
 - Web is not necessarily balkanized into distinct groups

Some High-level Properties of the Graph

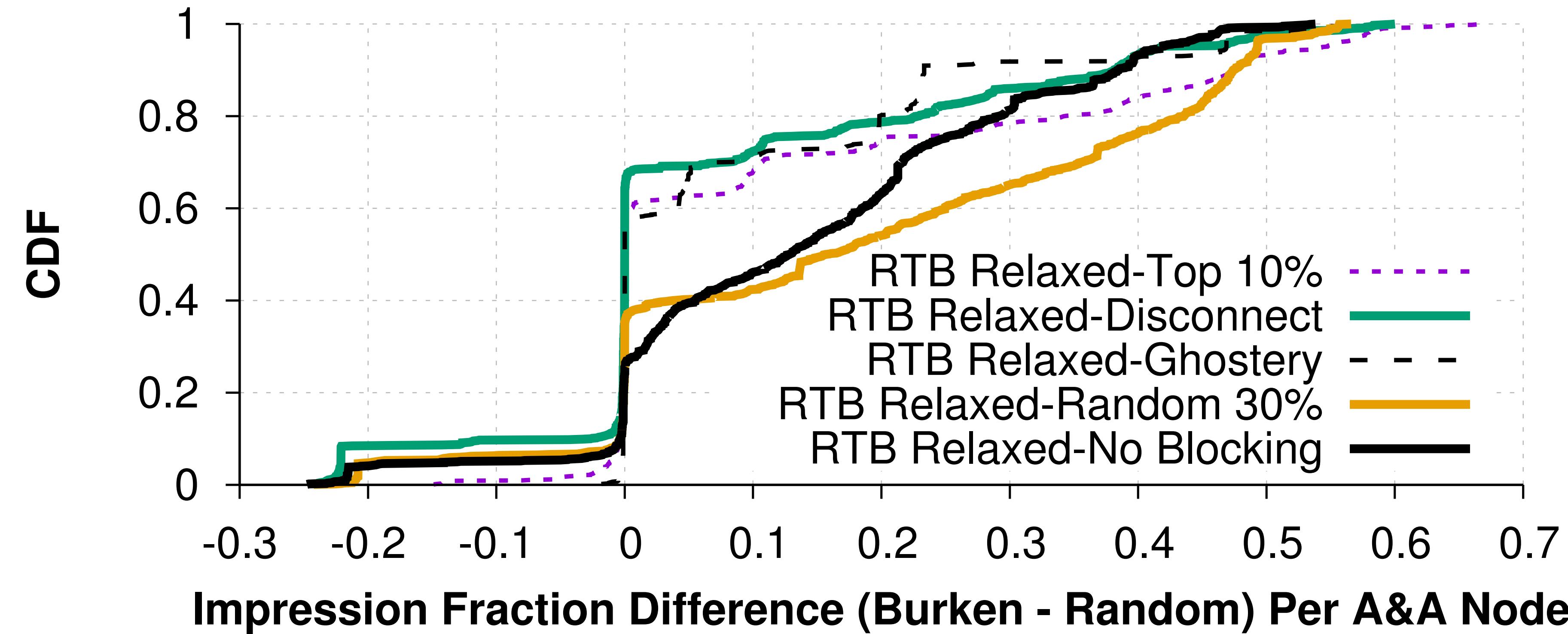
- The graph is very dense.
- No distinct communities
 - Web is not necessarily balkanized into distinct groups
- Expected top nodes with PageRank and Betweenness Centrality



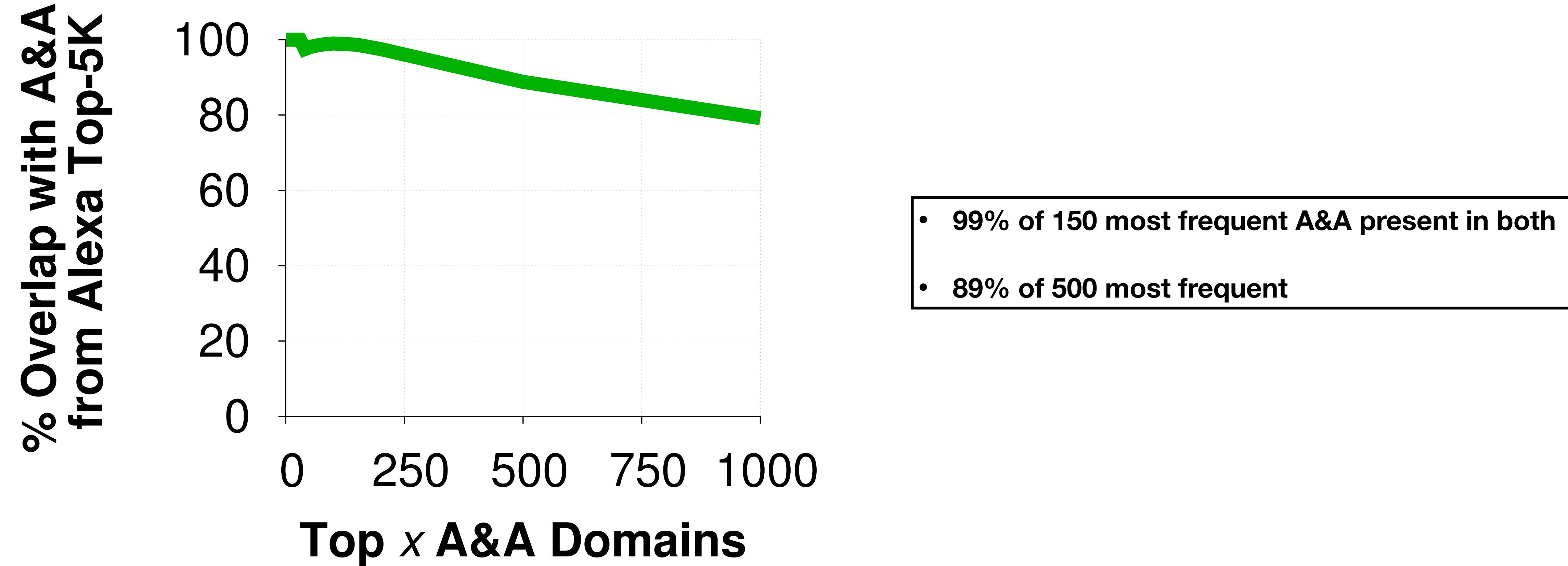
Simulations - Model Validation



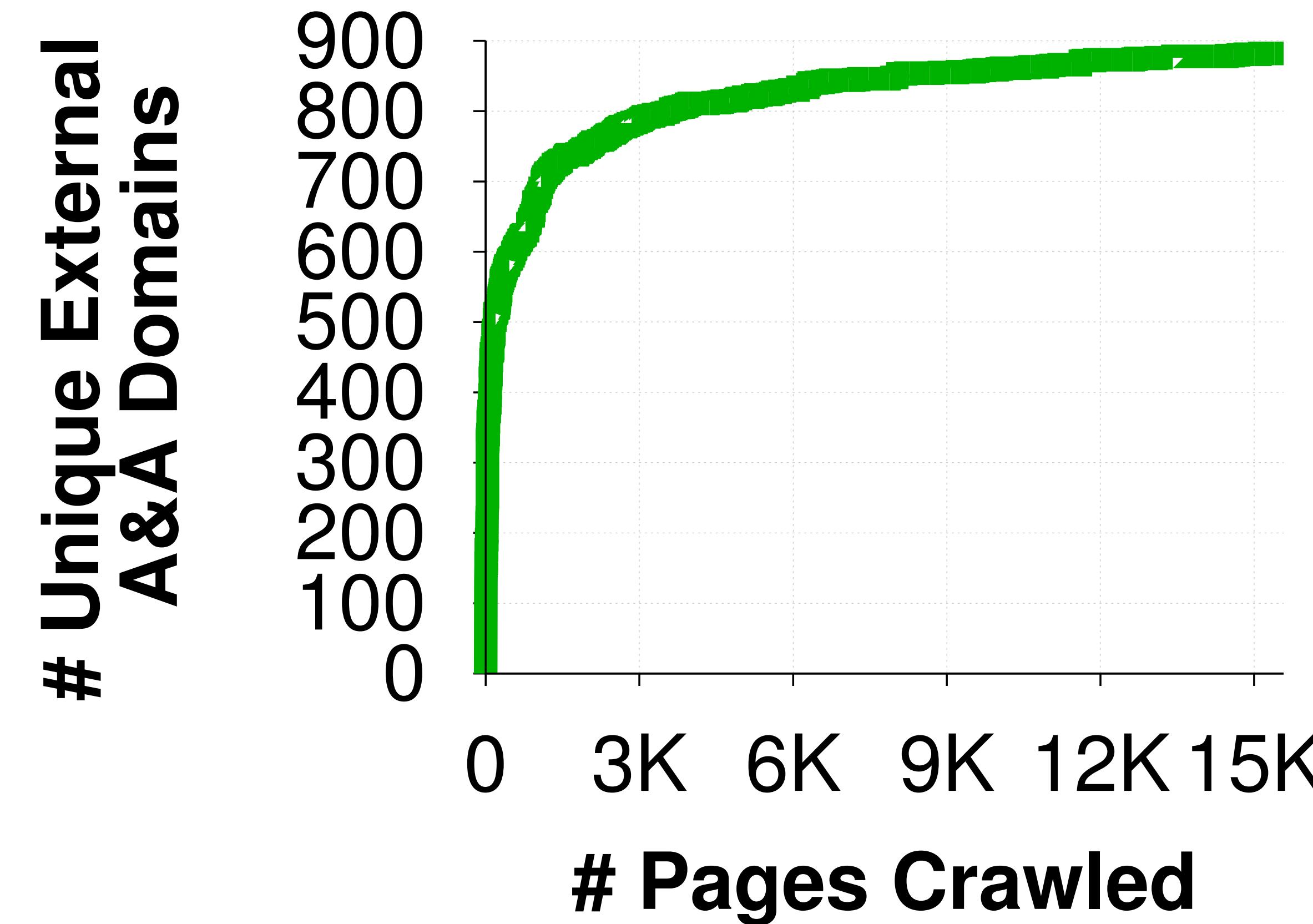
Simulations - Random Model



A&A Coverage

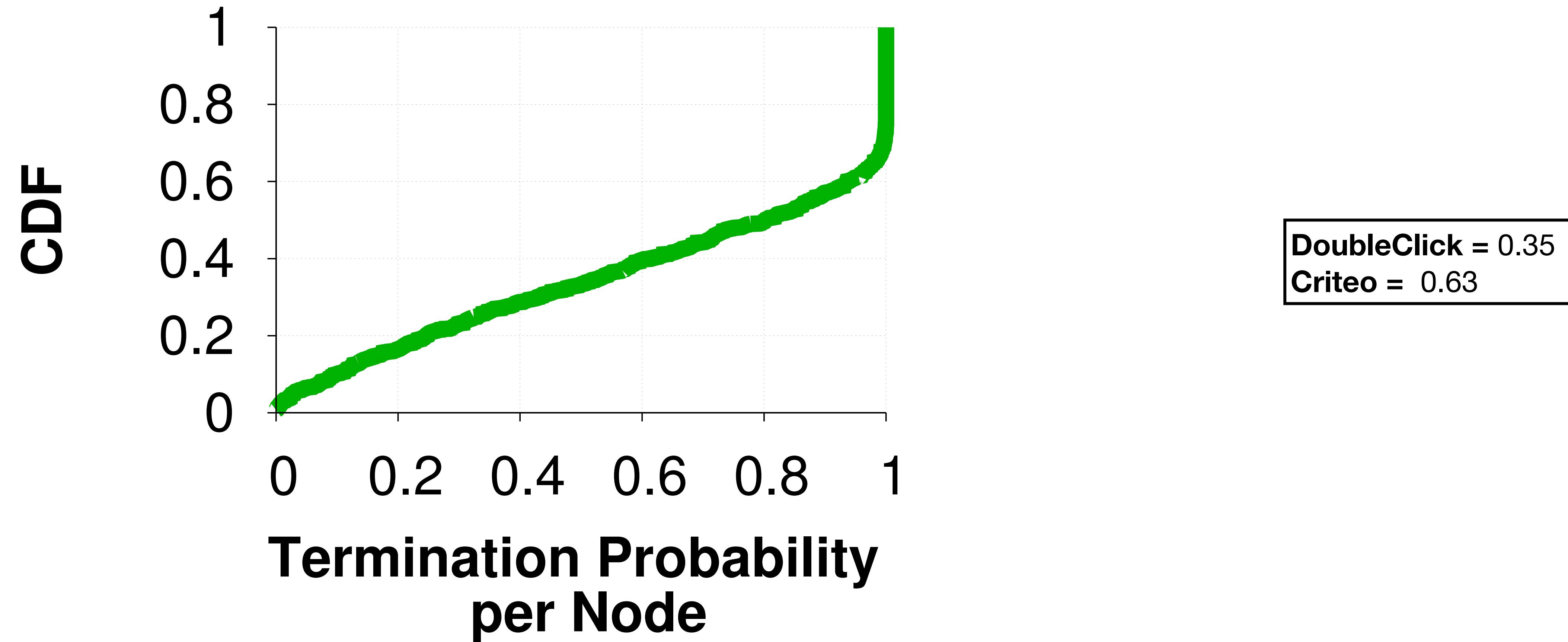


A&A Coverage (2)

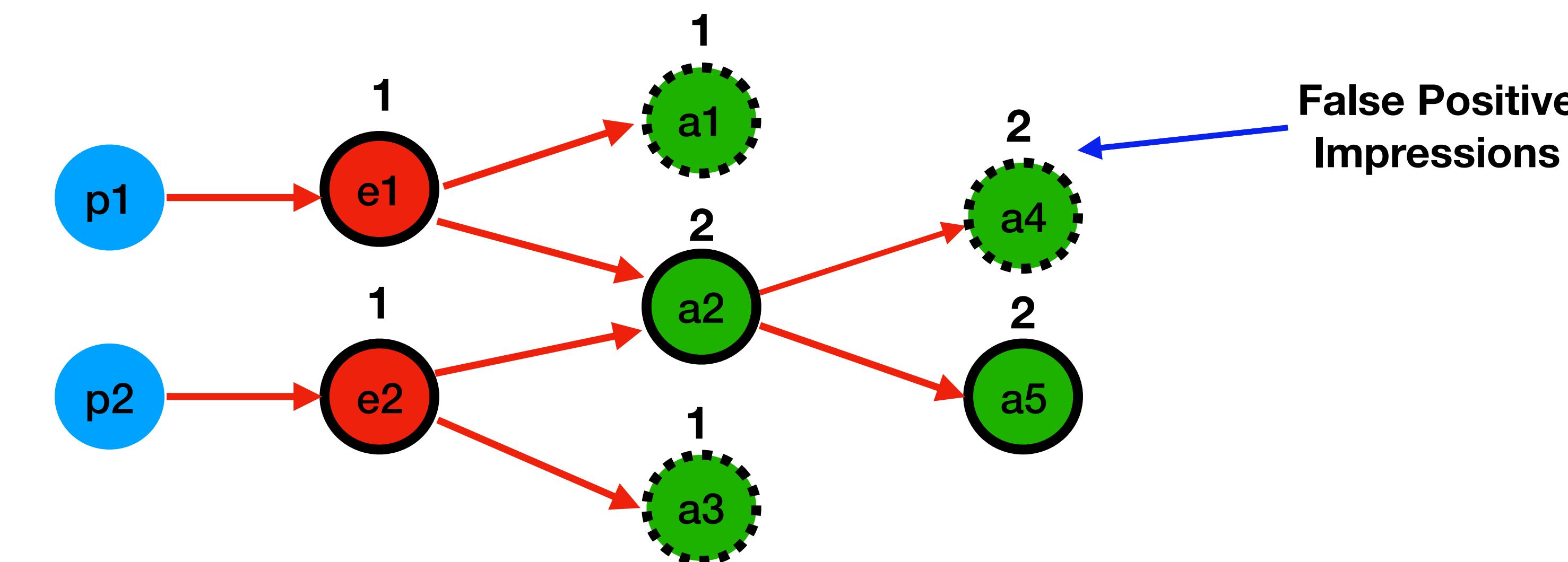
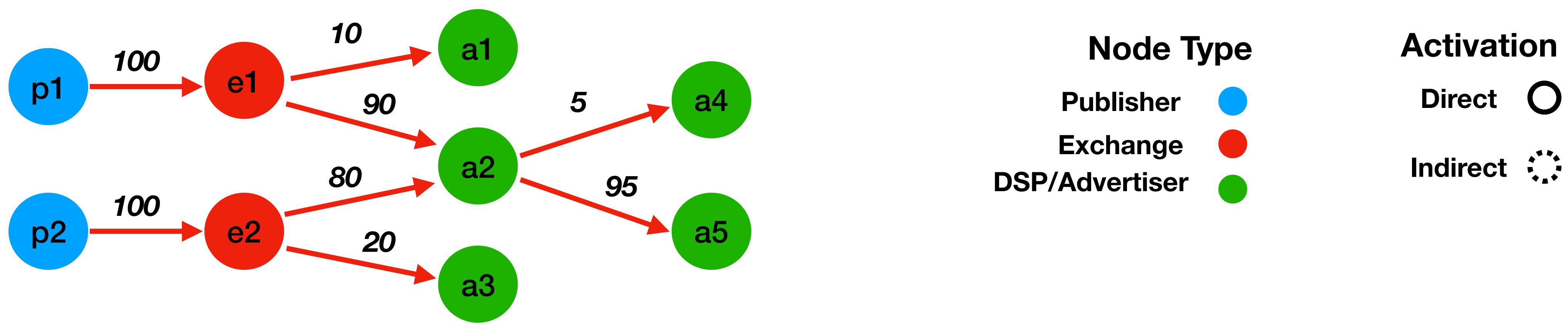


Number of unique A&A contacted by A&A domains as our crawl progresses

Termination Probability

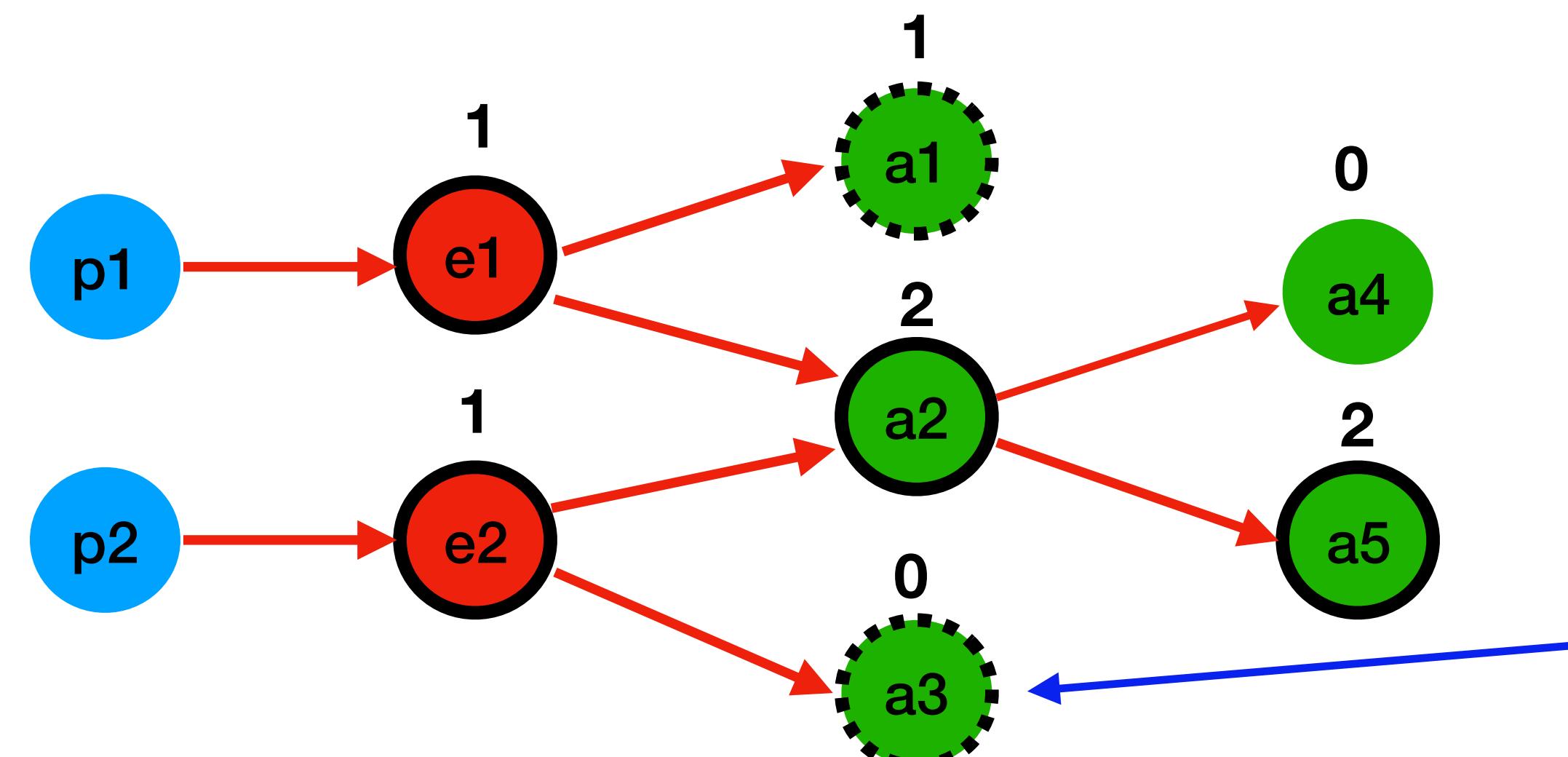
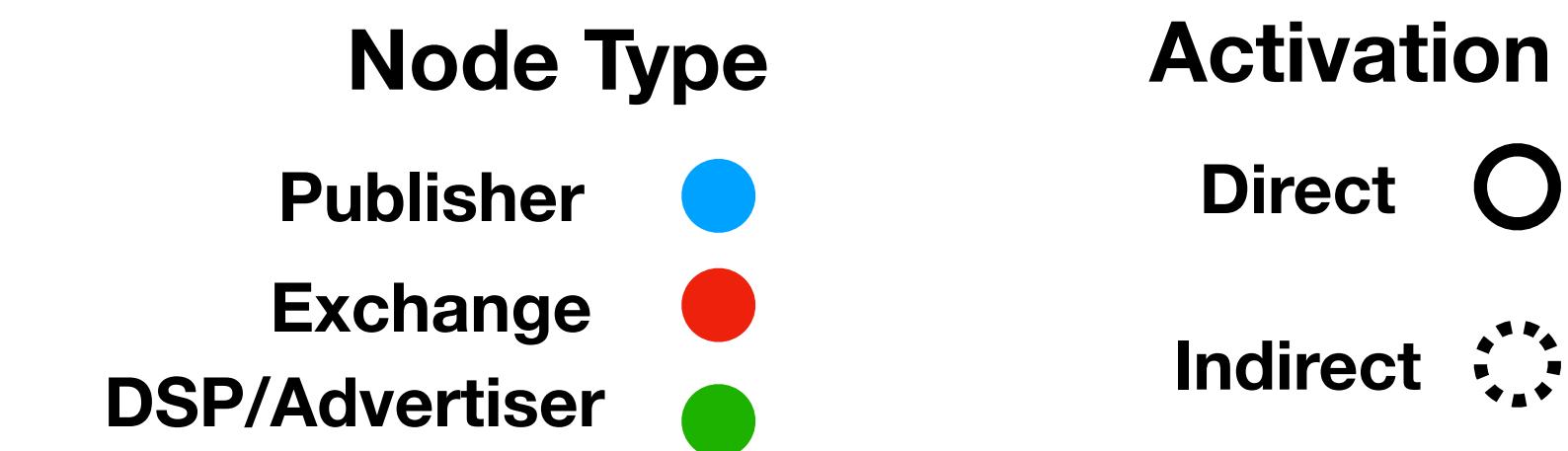
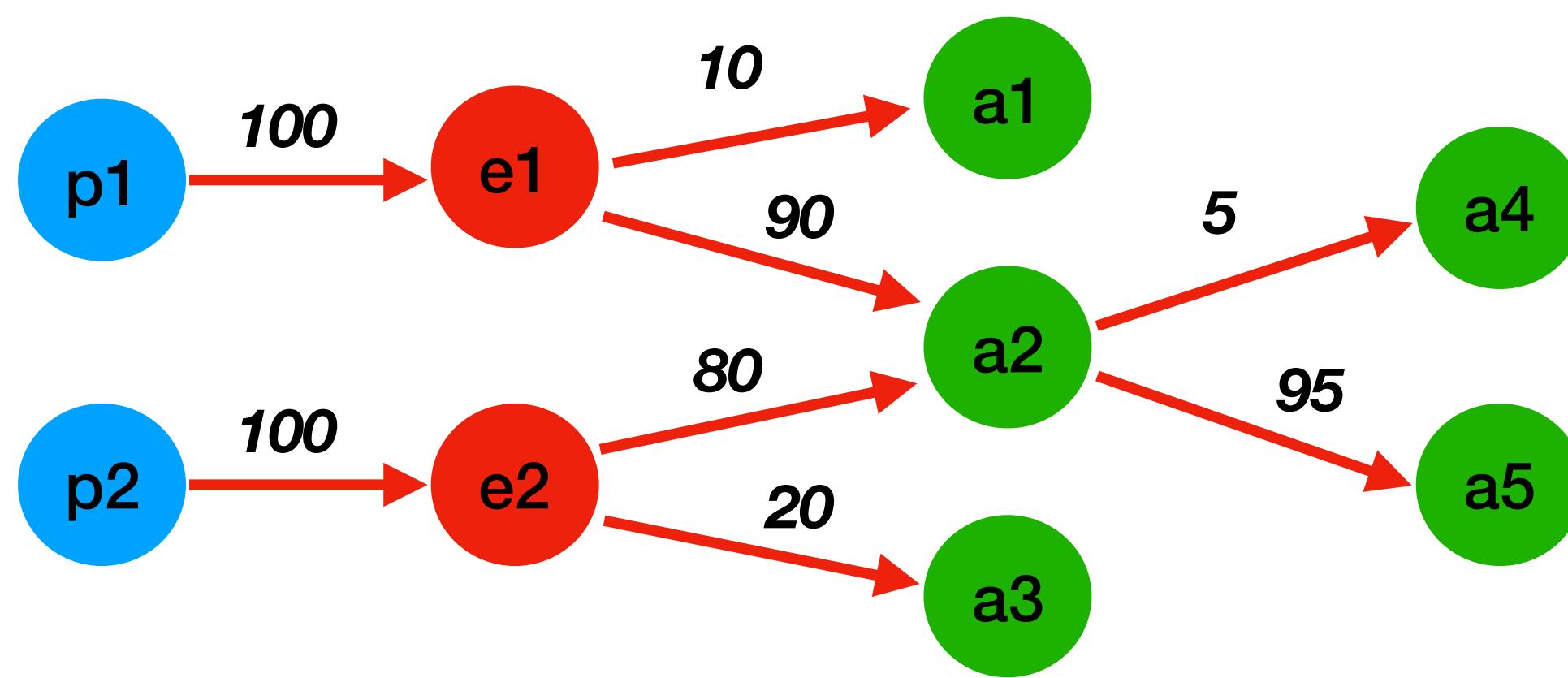


Simulation RTB Relaxed



Every A&A domain
acts as Ad Exchange

Simulation Cookie Matching Only



No Ad Exchange.
Indirect propagation only
along 200 CM edges

False Negative
Impressions